



Building a Longitudinal National Integrated Cardiovascular Database

— Lessons Learnt From SingCLOUD —

Khung Keong Yeo; Hean-Yee Ong; Terrance Chua; Zheng Jie Lim;
Jonathan Yap; Hee Hwa Ho; Fazlur Jaufferally, BSc; Khim-Leng Tong;
Pipin Kojodjojo, PhD; Hwee-Bee Wong, PhD; Derrick Heng; Kelvin Bryan Tan, PhD;
Arthur Mark Richards, MD, PhD; Kristine Leok-Kheng Teoh; Kenny Sin;
Ngiap Chuan Tan; Simon Biing Ming Lee; Terence Lim, BSc; Andy Ta;
Edwin Liok, BSc; Yee How Lau, BSc; Fei Gao, PhD; Christian Liman, BSc;
Joydeep Sarkar; Anders Sahlén, MD, PhD; Tian Hai Koh; Mark Y. Chan, PhD

Background: Real world data on clinical outcomes and quality of care for patients with coronary artery disease (CAD) are fragmented. We describe the rationale and design of the Singapore Cardiovascular Longitudinal Outcomes Database (SingCLOUD).

Methods and Results: We designed a health data grid to integrate clinical, administrative, laboratory, procedural, prescription and financial data from all public-funded hospitals and primary care clinics, which provide 80% of health care in Singapore. Here, we explain our approach to harmonize real-world data from diverse electronic medical and non-medical platforms to develop a robust and longitudinal dataset. We present pilot data on patients with myocardial infarction (MI) treated with percutaneous coronary intervention (PCI) between 2012 and 2014. The initial data set had 53,395 patients. Of these, 35,203 had CAD confirmed on coronary angiography, of whom 21,521 had PCI. Eventually, limiting to 2012–2014, 3,819 patients had MI with PCI, while 5,989 had MI. Compared with the quality improvement registry, Singapore Cardiac Data Bank, which had 189 fields for analysis, the SingCLOUD platform generated an additional 313 additional data fields, and was able to identify an additional 250 heart failure events, 664 major adverse cardiovascular events at 2 years, and low-density lipoprotein levels to 1 year for 3,747 patients.

Conclusions: By integrating multiple incongruent data sources, SINGCLOUD enables in-depth analysis of real-world cardiovascular “big data”.

Key Words: Big data; Database; Heart failure; Myocardial infarction

Cardiovascular diseases remain the main cause of mortality in most Organisation for Economic Co-operation and Development (OECD) countries, accounting for nearly one-third (32.3%) of all deaths in 2013.¹ Coronary artery disease (CAD) and heart failure

(HF) represent a spectrum of cardiovascular diseases, and are chronic illnesses that extend throughout each patient’s lifetime. National cardiovascular registries such as the American College of Cardiology (ACC) National Cardiovascular Data Registry (NCDR) and Sweden’s SwedeHeart

Received October 23, 2019; accepted October 23, 2019; J-STAGE Advance Publication released online December 7, 2019 Time for primary review: 1 day

Department of Cardiology (K.K.Y., T.C., J.Y., A.S., T.H.K.), Department of Cardiothoracic Surgery (K.S.), Singapore Cardiac Data Bank (Y.H.L.), NHRIS (F.G.), National Heart Centre Singapore; Duke-NUS Medical School (K.K.Y., T.C., F.J.); Department of Cardiology, Khoo Teck Puat Hospital (H.-Y.O.), Singapore; Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria (Z.J.L.), Australia; Department of Cardiology, Tan Tock Seng Hospital (H.H.H.); Department of Internal Medicine, Singapore General Hospital (F.J.); Department of Cardiology, Changi General Hospital (K.-L.T.); Division of Cardiology, Department of Medicine, Jurong Health (P.K.); Biostatistics and Research Branch (H.-B.W.), Public Health Group (D.H.), Policy Research and Evaluation Division (K.B.T.), Ministry of Health; Cardiovascular Research Institute [CVRI], National University Heart Centre [NUHCS], National University Health System (A.M.R.); Department of Cardiac, Thoracic and Vascular Surgery (K.L.-K.T.), Cardiac Department (M.Y.C.), National University Hospital; SingHealth Polyclinics (N.C.T.); National Healthcare Group Polyclinics (S.B.M.L.); Integrated Health Informatics System, Ministry of Health Holdings (T.L., A.T.); Infocomm Media Development Authority of Singapore (E.L.); Holmusk (C.L., J.S.), Singapore; and Karolinska Institute, Stockholm (A.S.), Sweden

ClinicalTrials.gov ID: NCT03760705

Mailing address: Khung-Keong Yeo, MBBS (Singapore), FAMS, FACC, FSCAI, ABIM (Internal Med, Cardiology & Interventional Cardiology, USA), National Heart Centre Singapore, 5 Hospital Drive, Singapore 169609. E-mail: yeo.khung.keong@singhealth.com.sg

ISSN-2434-0790 All rights are reserved to the Japanese Circulation Society. For permissions, please e-mail: cr@j-circ.or.jp

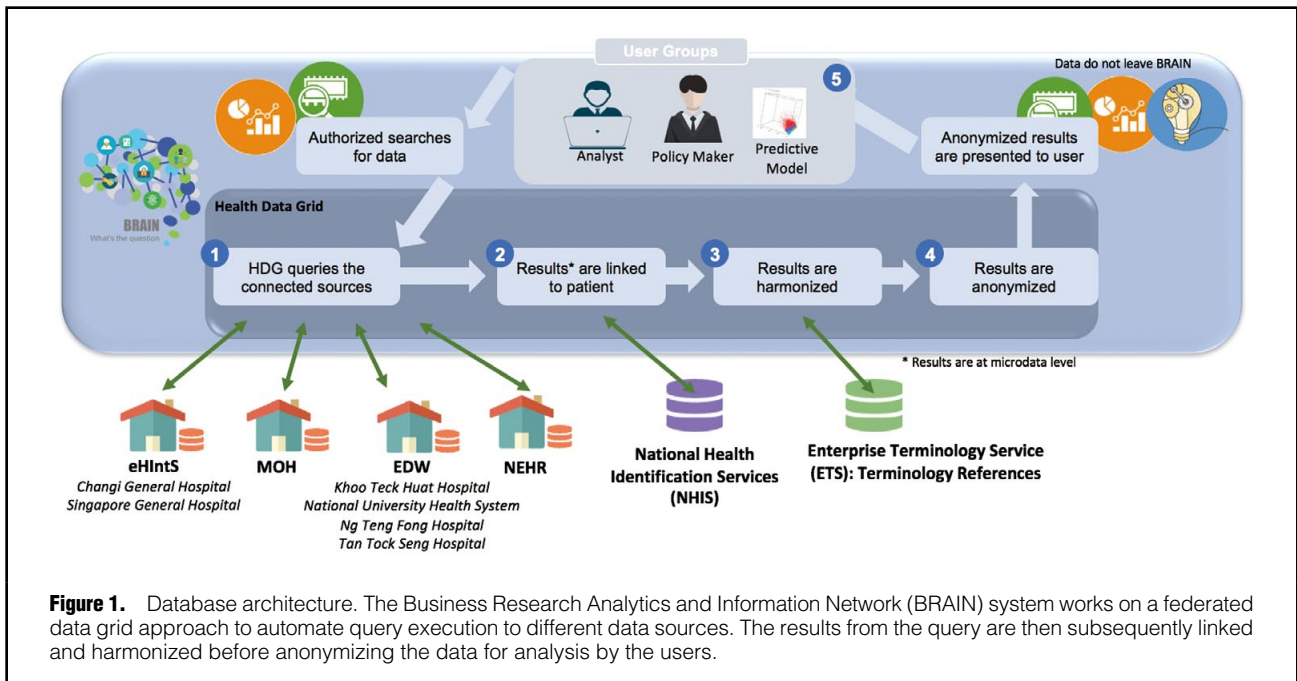


Figure 1. Database architecture. The Business Research Analytics and Information Network (BRAIN) system works on a federated data grid approach to automate query execution to different data sources. The results from the query are then subsequently linked and harmonized before anonymizing the data for analysis by the users.

have provided the cardiovascular community with important real-world insights.²⁻⁴ Limitations and challenges remain in the use of clinical registries, for example, in terms of data standardization, cost of implementation, and follow-up duration.⁵ We therefore sought to integrate real-world data from diverse data sources, both curated and un-curated, for patients with cardiovascular disease in Singapore. This paper describes the rationale and design of the observational Singapore Cardiovascular Longitudinal Outcomes Database (SingCLOUD) study.

Methods

Because of the constraints applied to the investigators in using the national level data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the correspondence author. Approved users may use the data but in a secure environment as mandated by the government.

Database Governance

The SingCLOUD program is led by a governance committee chaired by the principal investigator. Each hospital is represented by the chief of cardiology of the hospital or his/her appointee. The committee also includes a senior representative from the Ministry of Health (MOH). All executive decisions regarding any research projects or publications have to be unanimous. In addition, the Governance Committee appoints a recognized expert and researcher in cardiovascular medicine as a senior advisor. This senior advisor, however, does not have voting power. A publications subcommittee is also appointed to ensure the academic quality of research papers, as well as to adjudicate authorship matters.

SingCLOUD was approved with a waiver of written informed consent by the 2 centralized ethics review boards in Singapore: the SingHealth Central Institutional Review

Board and the National Healthcare Group Domain Specific Review Board. In order to facilitate the sharing of data and to define the terms of the project, a common research collaborative agreement between the principal investigators and chief executive officers of all 10 public health-care institutions, including the MOH, was signed on 18 August 2014. The study is registered with ClinicalTrials.gov, ID: NCT03760705.

Database Architecture

This ongoing database includes Singaporeans and permanent residents who attended any of the public hospitals ($n=5$) as adult cardiac patients, and the public outpatient clinics ($n=18$ in 2014) in Singapore, and excludes hospitals primarily for obstetric and gynecological, psychiatric and pediatric problems. Cases are identified in 2 ways: (1) source data from the Singapore Cardiac Data Bank (SCDB), a quality improvement cardiac registry tracking cardiac procedures, surgery and HF admissions in public hospitals; and (2) discharge diagnosis coding based on the Australian version of the ICD-9 (1999–1 January 2012) and ICD-10 (after January 2012) coding. For ICD-9, the codes are the 410–414 series, and for ICD-10, the codes are the I20–I25 series. The inclusion criterion was an inpatient or outpatient encounter with a health-care practitioner for the following proven or suspected diagnoses, associated with CAD, such as myocardial infarction (MI), acute coronary syndrome, stable angina and so on; or associated with congestive HF (CHF; e.g., systolic HF, diastolic HF, HF with preserved ejection fraction [EF], HF with reduced EF, cardiomyopathy etc.). To achieve the greatest possible coverage, we expanded the diagnostic search criteria to presentation with chest pain, sudden cardiac arrest, ventricular fibrillation or other tachyarrhythmias in which CAD is suspected.

Construction of the database required multiple electronic data sources including (1) clinical and procedural data collected by the SCDB; (2) clinical, administrative, pharmacy, financial data collected by each participating health-

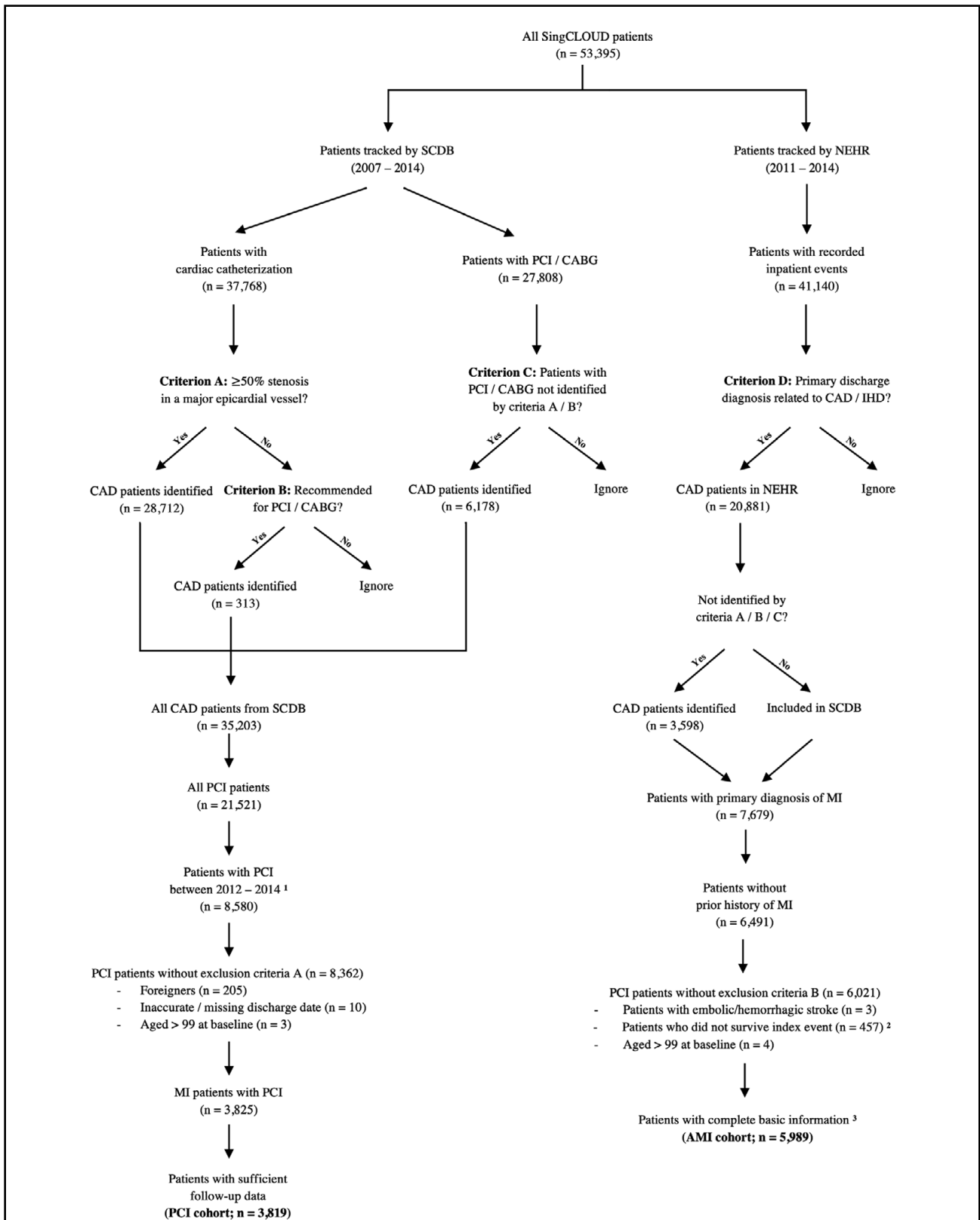


Figure 2. CONSORT diagram showing how Singapore Cardiovascular Longitudinal Outcomes Database (SingCLOUD) was used to generate the myocardial infarction (MI) and the MI with percutaneous coronary intervention (PCI) cohorts. AMI, acute myocardial infarction; CABG, coronary artery bypass grafting; CAD, coronary artery disease; IHD, ischemic heart disease; NEHR, National Electronic Health Record System; SCDB, Singapore Cardiac Data Bank. ¹Analysis was focused on 2012–2014 patients because the data were relatively more complete and had been audited internally. ²Removed patients who died ≤14 days after index MI. ³Demographic information such as gender, age, and ethnicity; the cohort has an overlap of 4,844 patients with SCDB records.

Table 1. Data Fields Audited in SingCLOUD

Subject area	Section	Data fields
Clinical data	Index admission	- Demographics - Primary diagnosis codes - Admission and discharge dates
	12-month readmissions	- Institution - Primary diagnosis codes - Admission and discharge dates
	Laboratory data (Index, 6 months and 12 months)	- Assessment date/time - Institution - Value and unit
	Medication (Index, 6 months and 12 months)	- Prescription date - Institution - Medication name, dosage and duration
Inpatient financial data	Index admission individual services	- Code and description - Gross, subsidy and net costs
	Index admission total cost	- Total gross, subsidy and net costs for admission
Outpatient financial data	12-month post-discharge total cost	- Total gross, subsidy and net costs for admission

SingCLOUD, Singapore Cardiovascular Longitudinal Outcomes Database.

care institution using their operational systems; and (3) administrative data collected by the MOH. The data from various sources flow to 6 main data warehouses before entering SingCLOUD: (i) eHINTS; (ii) EDW (both of which are data warehouses serving public hospitals); (iii) National Health Identification System (NHIS); (iv) Enterprise Terminology System (ETS); (v) MOH data hub; and (vi) National Electronic Health Record System (NEHR). These sources are linked in a health data grid using a Ministry platform called Business Research Analytics and Information Network (BRAIN). BRAIN and the individual electronic data sources are shown in **Figure 1** and are further described in **Supplementary File**.

Health Data Grid

The core concept of the platform is the utilization of a health data grid. Instead of duplication and repeated storage of information, the platform allows data extraction and analytics from multiple hospitals using the patient's National Registration Identity Card (NRIC) number as the core identifier. By law, under the National Registration Act of 1965, all Singaporean citizens and residents are required to have a Singapore NRIC, with a unique identification number. This unique NRIC number is used throughout Singapore's public health-care institutions, which serve the vast majority of Singapore's patient population. The NRIC serves as a nationwide means of patient identification across all electronic medical record systems. The NRIC is also used in all government services for administrative claims and the recording of births and deaths.⁶ Use of the NRIC is similarly ubiquitous in private sector services. While many data elements are common across health-care institutions such as laboratory data or discharge codes, the ETS system further ensures that commonly used terms are harmonized across the country. For specialty-specific terminology such as those found in cardiac catheterization or echocardiography reports, the SingCLOUD investigators further defined terminology to allow common mapping of common and important fields (e.g., left ventricular EF, coronary anatomy etc.). The logic for these efforts in data harmonization resides within

SingCLOUD.

Database Security and Personal Data Protection

Multiple measures have been put in place to prevent unauthorized access to SingCLOUD. In order to ensure patient confidentiality and anonymity, the data are then anonymized before being presented to the researcher. The system relies on dual-key encryption so that users at any institution cannot re-identify patients. The system that carries out these tasks is hosted on Ministry servers. For public health reasons, the Ministry has the authority to re-identify patients but this is governed by established approval processes. Access to anonymized data is possible only with the 2-factor identification verification of approved analysts, and analysis can be performed only on computers on the hospital network, via a virtual private network.

Proof-of-Principle Analysis Using SingCLOUD

The full dataset contains data from 2007 to 2014, but 1 of the key datasets (NEHR) was fully deployed to all public institutions only after 2011. Hence the pilot analysis was limited to patients admitted from 2012 to 2014. To test the robustness of the SingCLOUD health data grid, we focused our initial analysis on 2 groups of patients: (1) those who were admitted with an acute MI and who underwent percutaneous coronary intervention (PCI); and (2) those who were admitted with an MI (regardless of whether they had a PCI) at the 2 national cardiac centers in Singapore (**Figure 2**). These narrow subsets of patients were chosen because they represent the best characterized group of patients in SingCLOUD and because the cardiac catheterization and PCI data had already been audited clinically via the SCDB. The specific aims of the SingCLOUD project are as follows: (1) to establish the in-hospital and long-term clinical outcomes for death, stroke, and MI; (2) to identify the risk factors for major adverse cardiovascular events (MACE) including hospital re-admission; (3) to describe the quality-of-care provided in terms of physician adherence to guideline-recommended therapy (e.g., use of antiplatelet agents and statins); (4) to establish the overall cost of care including the in-hospital costs and the costs of outpatient

and subsequent episodes of care up to 3 years; and (5) to describe the referral pattern to and from the outpatient clinics. For the purpose of the pilot, we have focused only on aims (1–3).

Data Quality

For the data quality audit, the SingCLOUD data will be compared directly against the data from the source databases. The overall agreement rate, data completeness, and data accuracy for each domain will be calculated. Agreement rates will be calculated by number of correct matches divided by the number of fields audited. Two key subject areas of data will be audited: (1) clinical data; and (2) financial data. A brief description is provided in **Table 1**. Specific approval for this audit was obtained from the MOH.

The audited clinical data included index admission data from 1 center and follow-up data up to 12 months after the index admission from all contributing hospitals and polyclinics. Laboratory tests across 3 time periods (index admission, 6 months after index discharge, and 12 months after index discharge) were included. The specific tests were creatinine, CKMB, troponin T, HbA1C, hemoglobin, low-density lipoprotein (LDL), high-density lipoprotein, triglyceride and total cholesterol. Medication prescriptions were also reviewed across the same 3 time periods and included prescriptions for angiotensin-converting enzyme

inhibitors (ACEI), angiotensin receptor blockers (ARB), β -blockers, statins and anti-platelet drugs. The clinical audit data were manually collected by a team including a clinician through individual record review. For the financial data, the inpatient financial data included billing data during the admission. The outpatient financial data included billing data from all contributing hospitals and polyclinics, restricted to the 12 months after the index admission discharge date.

Comparisons between datasets to assess overall agreement rate, data completeness, and data accuracy for each subject area of index admission demographics, diagnosis and mortality, readmissions, medication, laboratory results, inpatient financial billing and outpatient financial billing were calculated. Agreement rates were calculated by number of matches divided by the number of fields audited.

For the clinical data, each data field was compared between the audit dataset and SingCLOUD dataset. Fields that did not exactly match between the 2 datasets were counted as mismatches. The number of mismatches was used to determine the agreement rates for each data field and overall agreement rate for the clinical data. Data completeness was assessed by considering mismatches due to missing data only in either dataset. Data accuracy was assessed by whether the actual value in the fields of both datasets matched. For the financial data comparison, agreement rates for each data field and overall agreement rate

Table 2. (A) Demographics and Baseline Patient Characteristics, (B) Disease and Procedural Characteristics		
(A)	PCI cohort (n=3,819)	MI cohort (n=5,989)
Demographics		
Gender		
Male	3,067 (80.3)	4,412 (73.7)
Female	752 (19.7)	1,577 (26.3)
Age (years)		
Mean \pm SD	60.5 \pm 11.7	63.1 \pm 13.1
Median (IQR)	60.0 (17.0)	69.0 (19.0)
<65 years	2,450 (64.2)	3,401 (56.8)
\geq 65 years	1,369 (35.8)	2,588 (43.2)
Race		
Chinese	2,410 (63.1)	3,910 (65.3)
Malay	662 (17.3)	1,053 (17.6)
Indian	591 (15.5)	771 (12.9)
Others	156 (4.1)	255 (4.3)
Medical history and risk factors		
Smoking history	1,736 (45.5)	–
Family history of CAD	298 (7.8)	439 (7.3)
Hypertension	2,365 (61.9)	2,929 (48.9)
Diabetes mellitus	1,362 (35.7)	2,251 (37.6)
Renal failure currently on dialysis	172 (4.5)	118 (2.0)
Cerebrovascular disease	225 (5.9)	–
Peripheral artery disease	103 (2.7)	–
Chronic lung disease	51 (1.3)	–
Hyperlipidemia	2,220 (58.1)	–
Prior CABG	224 (5.9)	76 (1.3)
Prior PCI	655 (17.2)	196 (3.3)
Prior MI	725 (19.0)	0 (0.0)
Prior HF	–	598 (10.0)

(Table 2 continued the next page.)

(B)	PCI cohort (n=3,819), n (%)	MI cohort (n=5,989), n (%)
MI classification		
STEMI	–	1,829 (30.5)
NSTEMI	–	3,620 (60.4)
Unknown	–	540 (9.0)
Type of procedure		
PCI†		3,377 (56.4)
Primary PCI for STEMI	1,844 (48.3)	
PCI for other MI	1,975 (51.7)	
CABG only	0 (0.0)	60 (1.0)
No procedure	0 (0.0)	2,552 (42.6)
CAD classification		
Diseased vessels‡		
SVD	1,443 (37.8)	1,477 (24.7)
DVD	1,070 (28.0)	1,221 (20.4)
TVD	1,084 (28.4)	1,567 (26.2)
Normal or minor CAD	165 (4.3)	307 (5.1)
Unknown	57 (1.5)	1,422 (23.7)
Left main disease	263 (6.9)	356 (5.9)
Baseline laboratory data and vital signs		
	Median (IQR)	Median (IQR)
Serum creatinine (μmol/L)	86.0 (34.0)	99 (42.0)
eGFR (mL/min/1.73 m ²)	82.6 (36.8)	78.1 (43.2)
Hemoglobin (g/dL)	14.3 (2.7)	13.8 (3.1)
Total cholesterol (mmol/L)	4.89 (1.67)	4.77 (1.67)
LDL (mmol/L)	3.02 (1.49)	2.94 (1.50)
HDL (mmol/L)	1.04 (0.34)	1.05 (0.36)
Triglycerides (mmol/L)	1.46 (1.04)	1.41 (0.99)
Ejection fraction (%)	54.0 (20.0)	50 (21.0)
SBP (mmHg)	125.0 (35.0)	–
DBP (mmHg)	71.0 (19.0)	–

(A) Data given as n (%), unless otherwise indicated. (B) †Includes patients who underwent both PCI and CABG. ‡Left main disease could be in conjunction with/without other diseased vessels.

CABG, coronary artery bypass graft; CAD, coronary artery disease; DBP, diastolic blood pressure; DVD, double-vessel disease; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; HF, heart failure; LDL, low-density lipoprotein; MI, myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction; PCI, percutaneous coronary intervention; SBP, systolic blood pressure; STEMI, ST-elevation myocardial infarction; SVD, single-vessel disease; TVD, triple-vessel disease.

for inpatient and outpatient financial data were calculated.

Statistical Analysis

The SingCLOUD platform allows for data analysis on the system platform itself. To ensure integrity and security of the system, data may not be extracted for analysis outside the platform. In order to facilitate this, different analytical tools and software were layered onto the system (i.e., SAS, STATA, SPSS, R, and Python), allowing different researchers to analyze the data using their preferred software. For the present analysis, Python 3.6.3 (Python Foundation) and Stata/MP 14.1 (StataCorp, College Station, TX, USA) were used.

Continuous variables are expressed as median (IQR) and categorical variables as n (%). One-year rates of mortality and MACE (defined as mortality, readmission for MI, or ischemic stroke/transient ischemic attack [TIA] events) were calculated for discharged patients for the 2 years after discharge. HF admissions were also calculated. Cox regression analysis was performed to model independent predictors of 1-year MACE. Only variables with $P < 0.01$ on univariate analysis included in the multivariate Cox

regression. We did not perform comparisons between the 2 cohorts because 1 cohort is a subset of the other, and this was not the purpose of the paper.

Physician prescription adherence (PPA) was analyzed at discharge and at 1 year from date of discharge for patients between 2012 and 2013 who survived index hospitalization. PPA was defined as present at 1 year after discharge if he/she prescribed for $\geq 80\%$ of the time during the 1 year after discharge. Patients who passed away before 1 year had the PPA rates calculated as a percentage of the time they were alive after discharge. These data were not limited to data from the admitting hospital; instead, these data track prescriptions for individual patients from all the public hospitals and outpatient clinics in Singapore. Again, readmissions were obtained using the SingCLOUD platform from across the entire country. For the purpose of this initial analysis, we did not examine the relationship of PPA adherence at other time points to 1-year mortality. The dataset does not include prescriptions filled by private pharmacies or private clinics. Public institutions, however, generally provide substantial subsidies, and patients who continue to be managed in public institutions will obtain

Table 3. (A) Long-Term Clinical Outcomes, (B) Physician Prescription Adherence, (C) MACE According to PPA at Discharge (2012–2013)				
(A)	PCI cohort (n=3,819)		MI cohort (n=5,989)	
	1 year	2 years	1 year	2 years
Post-discharge mortality [†]	128 (3.4)	176 (4.6)	580 (9.7)	770 (12.9)
Total mortality [‡]	331 (8.7)	379 (9.9)	–	–
MI recurrence	301 (7.9)	376 (9.8)	750 (12.5)	908 (15.2)
Stroke/TIA	62 (1.6)	79 (2.1)	66 (1.1)	91 (1.5)
HF	216 (5.7)	250 (6.5)	299 (5.0)	377 (6.3)
MACE	426 (11.2)	528 (13.8)	1,089 (18.2)	1,349 (22.5)
MACE with HF	561 (14.7)	664 (17.4)	1,260 (21.0)	1,511 (25.2)
(B)	PCI cohort (n=3,616[†])		MI cohort (n=5,989)	
	Discharge	1 year	Discharge	1 year
Aspirin	3,450 (95.4)	2,843 (78.6)	4,656 (77.7)	3,897 (65.1)
P2Y12 inhibitors [‡]	3,513 (97.2)	2,392 (66.2)	4,649 (77.6)	2,888 (48.2)
DAPT [§]	3,425 (94.7)	2,176 (60.2)	4,380 (73.1)	2,383 (39.8)
ACEI	2,255 (62.4)	1,605 (44.4)	3,020 (50.4)	2,146 (35.8)
ARB	438 (12.1)	533 (14.7)	701 (11.7)	662 (11.1)
ACEI/ARB	2,689 (74.4)	2,246 (62.1)	3,695 (61.7)	3,139 (52.4)
LLT (statins)	3,467 (95.9)	2,890 (79.9)	4,826 (80.6)	4,251 (71.0)
LLT (non-statins)	233 (6.4)	191 (5.3)	301 (5.0)	258 (4.3)
β-blockers	3,131 (86.6)	2,601 (71.9)	4,394 (73.4)	3,856 (64.4)
(C)	PPA patients	Non-PPA patients	P-value	
Aspirin	410 (17.4)	34 (27.2)	0.008	
P2Y12 inhibitors	425 (17.7)	19 (26.0)	0.086	
ACEI	245 (15.6)	199 (22.0)	<0.001	
ARB	75 (24.1)	369 (17.0)	0.003	
ACEI/ARB	319 (17.0)	125 (20.8)	0.038	
LLT (statins)	417 (17.6)	27 (26.0)	0.036	
LLT (non-statins)	37 (22.8)	407 (17.6)	0.111	
β-blockers	371 (17.3)	73 (21.8)	0.055	

(A) Data given as n (%). Outcomes of MI, ischemic stroke/TIA and HF do not include in-hospital complications or events. Given that the MI cohort excludes patients who died in hospital, total mortality is not calculated for the MI cohort. [†]Excludes patients who died in hospital; [‡]covers both in-hospital and post-discharge mortality. MACE, major adverse cardiovascular events (consisting of post-discharge mortality, recurrent MI, and ischemic stroke/TIA events; post-discharge HF events were not included unless otherwise indicated). (B) Data given as n (%). [†]Reduced due to exclusion of in-hospital mortality cases from the analysis. [‡]Clopidogrel, prasugrel, ticagrelor, and ticlopidine; [§]aspirin and P2Y12 inhibitors. (C) Data given as n (%). Only patients who survived index hospitalization were included in the analysis. MACE, major adverse cardiovascular events (consisting of all-cause mortality, recurrent MI, and ischemic stroke event, and onset of heart failure). ACEI, angiotensin-converting enzyme inhibitors; ARB, angiotensin receptor blockers; CVA, cerebrovascular accident; DAPT, dual antiplatelet therapy; HF, heart failure; LLT, lipid-lowering therapy; MI, myocardial infarction; PPA, physician prescription adherence; PCI, percutaneous coronary intervention; TIA, transient ischemic attack.

the medications from the public institutions. Regardless, because the study examines PPA, where the patient has the prescription filled would not be relevant in this analysis. We calculated PPA rates for guideline-directed drugs: statins, aspirin, P2Y12 inhibitors, ACEI/ARB, lipid-lowering therapy and β-blockers. The MACE rates between PPA and non-PPA patients were also compared, using Fisher's exact test.

Results

The initial pilot had 53,395 patients from 2007 to 2014. Of these, 35,203 had CAD confirmed on coronary angiography, of whom 21,521 had PCI. Limiting to the period 2012–2014, 3,819 patients were identified as having MI with PCI, while 5,989 were identified as having MI (Figure 2). Table 2A lists demographics and clinical characteristics for both cohorts. Table 2B lists the main laboratory, angiography and procedural data. Table 3A lists 1- and 2-year

post-discharge clinical outcomes including mortality, length of stay, MI recurrence, stroke/TIA, HF admissions and composite MACE. For the MI with PCI cohort, 1- and 2-year MACE were 11.2% and 13.8%, respectively. For the MI cohort, these were 18.2% and 22.5%, respectively. Table 3B lists PPA to guideline-recommended drugs at the time of discharge and at 1 year from the date of discharge. At 1 year, for the MI with PCI, and MI cohorts, respectively, PPA was 62.0% and 37.2% for dual-antiplatelet therapy; 86.0% and 61.7% for aspirin; and 88.2% and 67.3% for statins. Initial analysis examining the effect of PPA is shown in Table 3C, with univariate analysis of MACE events according to adherence to specific drugs. For example, in patients in whom there was PPA to statins, the MACE rate was 17.6% at 1 year compared with 26.0% in non-PPA patients.

Table 4A lists the intersection of unique patients from the different data sources using the MI with PCI cohort as an example. For the SingCLOUD Pilot, even though index

Table 4. (A) Intersection of Data With SCDB PCI Cohort, (B) MI With PCI Cohort Demographics (n=3,819) vs. Follow-up Status, (C) Clinical Outcomes: SingCLOUD vs. Single-Hospital Data			
(A)	Cohort name	No. unique patients	No. patients intersecting with PCI cohort
1	Echo	6,038	123
2	Mortality	9,864	393
3	Nuclear study	9,274	581
4	Demographics	54,248	3,819
5	Event and Diagnosis	48,928	3,819
6	Laboratory data	44,976	3,588
7	Medication	31,560	2,209
8	Inpatient-EDW	30,686	2,424
9	Inpatient-EHINTS	35,225	2,172
10	Outpatient-EDW	41,252	3,083
11	Outpatient-EHINTS	37,551	2,388
(B)	With follow-up (n=3,569)	Without follow-up (n=250)	P-value
Demographics			
Gender			<0.001
Male	2,892 (81.0)	175 (70.0)	
Female	677 (19.0)	75 (30.0)	
Age (years; numerical)			<0.001
Mean±SD	60.2±11.6	66.1±12.7	
Median (IQR)	60 (16)	66.0 (18.7)	
Age (years; categorical)			<0.001
<65	2,340 (65.6)	110 (44.0)	
≥65	1,229 (34.4)	140 (56.0)	
Race			0.525
Chinese	2,250 (63.0)	160 (64.0)	
Malay	613 (17.2)	49 (19.6)	
Indian	558 (15.6)	33 (13.2)	
Others	148 (4.1)	8 (3.2)	
(C)	PCI cohort (n=3,819)	Single-hospital PCI (n=1,098) [†]	
Post-discharge mortality [‡]	128 (3.4)	–	
Total mortality [§]	331 (8.7)	60 (5.5)	
MI recurrence	301 (7.9)	14 (1.3)	
Stroke/TIA	62 (1.6)	6 (0.5)	
CHF	216 (5.7)	–	
MACE	426 (11.2)	83 (7.6)	
MACE with CHF	561 (14.7)	–	

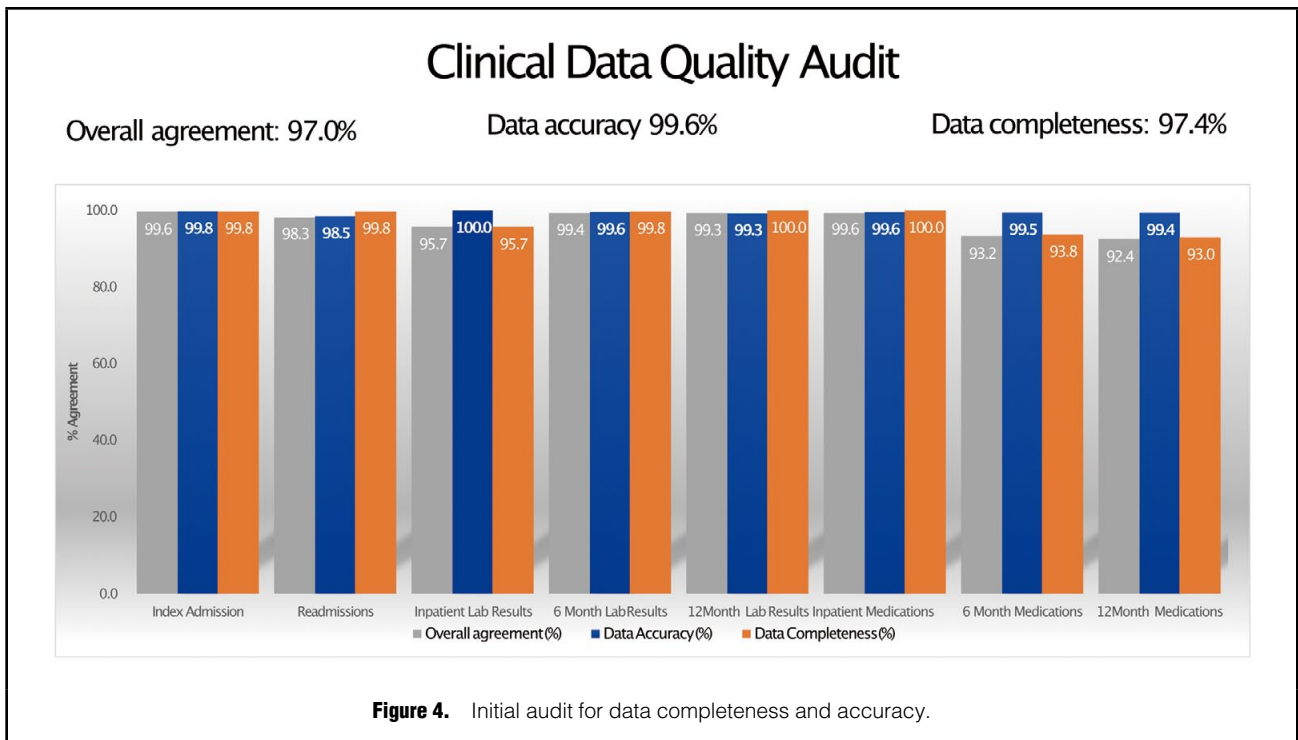
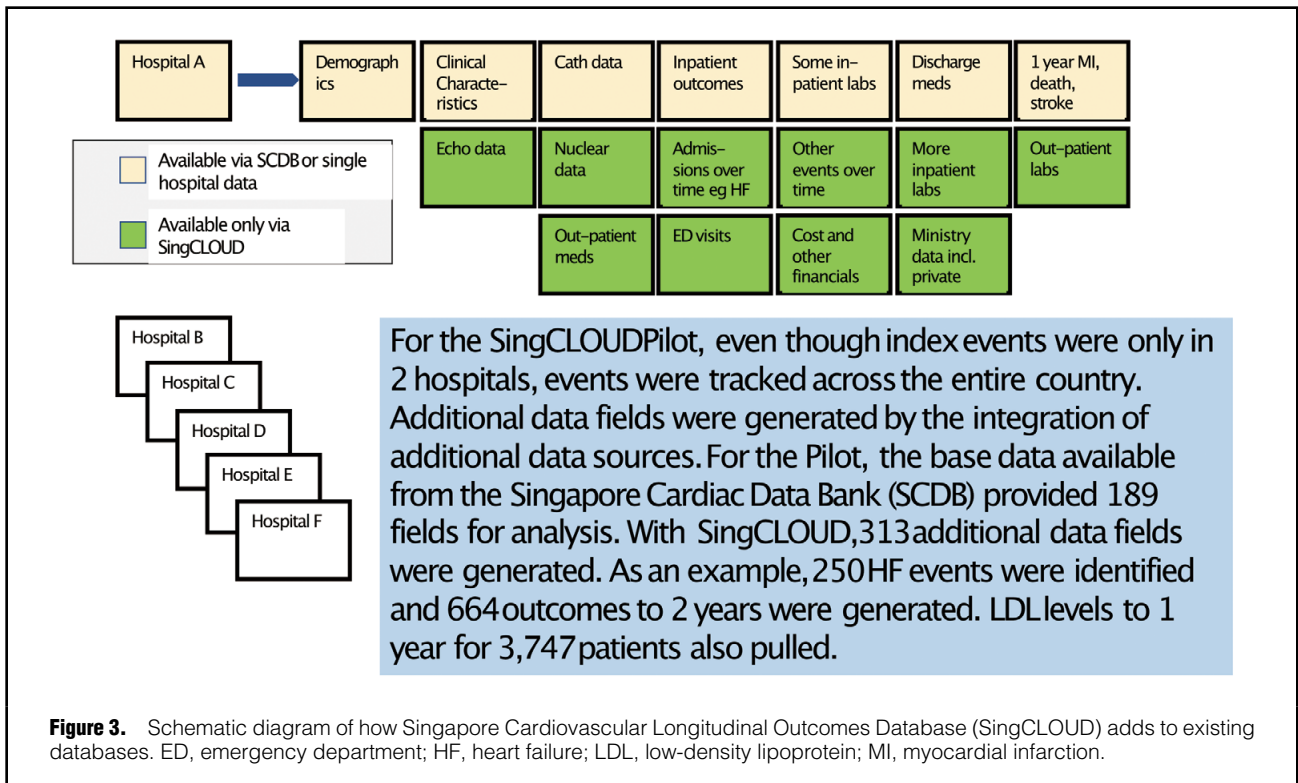
(A) No. PCI patients, 3,819. (B) Data given as n (%), unless otherwise indicated. (C) Data given as n (%). Outcomes of MI, ischemic stroke/TIA, and CHF do not include in-hospital complications or events. [†]Due to data constraints, single-hospital PCI data refer only to 30 days instead of 1 year. Post-discharge mortality for single-hospital PCI is defined differently and thus would be excluded. [‡]Excludes patients who died in hospital; [§]covers both in-hospital and post-discharge mortality. CHF, congestive heart failure; CVA, cerebrovascular accident; EDW, Enterprise Data Warehouse; EHINTS, Electronic Health Intelligence System; MACE, major adverse cardiac events (consisting of post-discharge mortality, MI recurrences, and ischemic stroke/TIA events; post-discharge CHF events were not included unless otherwise indicated); MI, myocardial infarction; PCI, percutaneous coronary intervention; SCDB, Singapore Cardiac Data Bank; SingCLOUD, Singapore Cardiovascular Longitudinal Outcomes Database; TIA, transient ischemic attack.

events were only in 2 hospitals, events were tracked across the entire country. Additional data fields were generated by the integration of additional data sources. For example, the base data available from SCDB provided 189 fields for analysis. With SingCLOUD, 313 additional data fields were generated. From this, 250 HF events were identified and 664 outcomes to 2 years were generated. LDL levels to 1 year for 3,747 patients were also identified (**Figure 3**).

Data Discrepancies

The data audit is ongoing, with 239 patients audited. **Figure 4** shows the summary audit data for clinical data. Overall agreement was 96.0%, with data completeness at 97.4%. For inpatient financial data, initial audit indicated a 97.0% overall agreement, with 100% agreement for total gross cost per patient.

A subgroup analysis was also performed to see if there were patients who did not appear in SingCLOUD when



tracked across the platform after the index admission. Using the MI with PCI cohort of 3,819 patients, 250 patients had zero encounters (i.e., no admissions, no mortality, no laboratory data and no medication). Their demographics are listed in **Table 4B**. With these patients, it is possible

that they are seeking medical care in the private sector and have no events that have resulted in mandated reporting to the Ministry (e.g., death or recurrent MI or stroke). Finally **Table 4C** lists the additional outcomes data gained when looking at the MI with PCI cohort using SingCLOUD

compared with outcomes from single-center data alone.

Discussion

The SingCLOUD program uses the combination of existing clinical and operational/administrative electronic data sources in a largely constrained patient population, in which the use of a national identification system allows for robust tracking of clinical events, to perform a comprehensive study of cardiovascular patients over time. The promise of such a system is that it allows for near complete tracking of events without the huge costs that would have otherwise been associated with formal tracking using clinical coordinators.

Challenges and Lessons Learnt

To establish the system, there were many key prerequisites. First, the clinician leadership at each hospital had to commit to work together. This ground-driven approach allowed the principal investigator to engage hospital administrative and Ministry leadership. Second, there was a willingness by the government to support this initiative; recognizing that such a platform was a key enabling technology for the use of “big data” and other data-driven technologies that can lower the cost of research studies. Third, the data grid platform allowed for data linkage and anonymization, without exposure of operational hospital systems to the risk of security/privacy breaches, loss of data or impact on operational system speed. Next, a robust governance policy that addressed ethics, publication, legal issues and the underlying philosophy of the program established clear rules for the investigators, including the Ministry itself. Last, we had to explore different mechanisms to fund the platform. This extended beyond government and hospital research grant support. We also explored public-private partnerships, in which private entities funded research supported by the Governance Committee.

Comparison With Other National Databases

In the proof-of-principle analysis using the SingCLOUD platform to track PPA and cardiovascular events in patients with MI, the use of guideline-directed medical therapy such as antiplatelet therapy and statins was consistent with international norms. The prescription of statins in the present group of MI patients with PCI was 88.2% at 1 year, compared with approximately 60% from a cohort study by Shah et al out of Olmsted County, USA.⁷ The Olmsted data were based on filled prescriptions by patients, while the present data were based on physician prescriptions data, which may account for some of these differences. In Olmsted County, there was linkage of all medical records from all sources of care through a centralized system, allowing for tracking of outcomes, similar to the present database. The present data are similar to those in a recent publication by Colantonio et al involving 29,932 Medicare beneficiaries admitted for MI.⁸ In that study, the 6-month discontinuation rate of statins was 12.3%.⁸ These data were obtained from the Medicare database using primarily Medicare pharmacy claims. In the present study, we have shown PPA only at 1 year, with 1-year PPA at 79.9% for the MI with PCI cohort. We expect that with the SingCLOUD platform, we will be able to study this over longer periods.

Next, clinical outcomes, such as recurrent MI at 9.3% for the MI with PCI cohort, are consistent with data from

contemporary published US data.⁹ The ACC NCDR relies on voluntary data submission by each hospital and combines these data in various analyses with data from other sources such as claims data from the Centers for Medicare and Medicaid.³ Although powerful due to the very high quality of data collected and the extensive research experience, the data are limited due to the submission from participating sites only. Also, for Medicare data, these would apply only to patients aged ≥ 65 years. Furthermore, outpatient data, including prescriptions and costs, are not captured. The Swedish platform SwedeHeart is an equally powerful and comprehensive platform utilizing a Web-based platform that allows the user to enter data directly into the system.¹⁰ It is integrated with the national death register, hospital admission data and prescription data, thereby allowing for a very comprehensive overview of heart disease in Sweden. It does not, however, have data from regular outpatient clinic encounters. In Asia, the Korean Acute Myocardial Infarction Registry (KAMIR) and the Japanese Acute Myocardial Infarction Registry (JAMIR) and JROAD registry are MI and cardiovascular registries, but they capture data only from contributing centers and do not track outpatient data in a systematic manner.^{11–13} In contrast, SingCLOUD extracts data from all public hospitals and from MOH, tracks outcomes across all public primary care sites clinics, and includes clinical, laboratory, prescription, administrative and financial data over time. Even data from private hospitals are captured if claims to the Ministry are submitted. This integration of data from multiple electronic sources enables a true longitudinal analysis of a patient's entire journey as they transition from the in-hospital period to primary care.

Study Limitations

Several limitations of SingCLOUD in its current iteration should be addressed. First, PPA does not necessarily indicate that the patient was actually taking the drug. Similarly, there are contextual data that may be relevant in studying laboratory or procedural data, which are currently not available in SingCLOUD. For example, these data do not capture reasons for not adhering to prescriptions for guideline-recommended medications, such as bleeding, allergies and so on. These disadvantages, however, are mitigated by the significantly larger sample size and the more complete capture of data. Second, the platform relies on a core nidus of clinically curated data from SCDB, which have been clinically audited. There remains the possibility that errors in data integration or harmonization may result in unintentional errors that are difficult to identify. To address this issue, an ongoing process of audit is underway and will include all segments of data including financial data. Although the initial audit findings have been excellent (**Figure 4**), the audit has also helped identify areas of data inaccuracy and incompleteness for which work is constantly ongoing to improve the database and data grid logic.

Conclusions

First, we believe that this platform demonstrates the use of technology to improve the delivery of quality and value-conscious health care to patients and to society. By reducing the cost of data collection, researchers and policy-makers can now study long-term outcomes, including medication prescription, laboratory data and administrative data,

without resorting to expensive cohort studies. Second, this platform is also a potent mechanism for use in cohort studies that may accrue events over a period of time. These cohort studies may be defined by clinical entities such as diabetes mellitus, or may even be genomic cohorts. Similarly, SingCLOUD can even be used as a pragmatic way to track clinical outcomes in medication or other intervention trials. Learning from the SingCLOUD program and continuously improving the BRAIN platform will only improve the accuracy and efficiency of the entire data ecosystem.

Acknowledgments

We acknowledge the hard work and contributions of the staff and database coordinators of the Singapore Cardiac Data Bank and SingCLOUD. The project is supported by the National Medical Research Council New Investigator Grant HSRNIG12nov012 (to K.K.Y.), and direct grants from the Ministry of Health and Infocomm and Media Development Authority of Singapore. Holmusk, a data science company, provided in-kind and funding support for the analysis of anonymized data. Holmusk received funding support from the Economic Development Board of Singapore, grant number COY-15-IDS-STD/17005.

Disclosures

C.L. and J.S. are employees of Holmusk. The other authors declare no conflicts of interest.

References

1. OECD. Health at a glance 2015: OECD indicators. http://dx.doi.org/10.1787/health_glance-2015-en (accessed September 1, 2019).
2. Alabas OA, Gale CP, Hall M, Rutherford MJ, Szummer K, Lawesson SS, et al. Sex differences in treatments, relative survival, and excess mortality following acute myocardial infarction: National cohort study using the SWEDEHEART Registry. *J Am Heart Assoc* 2017; **6**: e007123.
3. Masoudi FA, Ponirakis A, de Lemos JA, Jollis JG, Kremers M, Messenger JC, et al. Trends in U.S. cardiovascular care: 2016 report from 4 ACC national cardiovascular data registries. *J Am Coll Cardiol* 2017; **69**: 1427–1450.
4. Sahlen A, Varenhorst C, Lagerqvist B, Renlund H, Omerovic E, Erlinge D, et al. Outcomes in patients treated with ticagrelor or clopidogrel after acute myocardial infarction: Experiences from SWEDEHEART registry. *Eur Heart J* 2016; **37**: 3335–3342.
5. Bhatt DL, Drozda JP Jr, Shahian DM, Chan PS, Fonarow GC, Heidenreich PA, et al. ACC/AHA/STS statement on the future of registries and the performance measurement enterprise: A report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and The Society of Thoracic Surgeons. *J Am Coll Cardiol* 2015; **66**: 2230–2245.
6. Lee CE, Satku K. Singapore's health care system: What 50 years have achieved. New Jersey: World Scientific, 2016.
7. Shah ND, Dunlay SM, Ting HH, Montori VM, Thomas RJ, Wagie AE, et al. Long-term medication adherence after myocardial infarction: Experience of a community. *Am J Med* 2009; **122**: 961.e7–e13.
8. Colantonio LD, Huang L, Monda KL, Bittner V, Serban MC, Taylor B, et al. Adherence to high-intensity statins following a myocardial infarction hospitalization among medicare beneficiaries. *JAMA Cardiol* 2017; **2**: 890–895.
9. Chaudhry SI, Khan RF, Chen J, Dharmarajan K, Dodson JA, Masoudi FA, et al. National trends in recurrent AMI hospitalizations 1 year after acute myocardial infarction in Medicare beneficiaries: 1999–2010. *J Am Heart Assoc* 2014; **3**: e001197.
10. Jernberg T, Attebring MF, Hambraeus K, Ivert T, James S, Jeppsson A, et al. The Swedish Web-system for enhancement and development of evidence-based care in heart disease evaluated according to recommended therapies (SWEDEHEART). *Heart* 2010; **96**: 1617–1621.
11. Kim HK, Jeong MH, Lee SH, Sim DS, Hong YJ, Ahn Y, et al. The scientific achievements of the decades in Korean Acute Myocardial Infarction Registry. *Korean J Intern Med* 2014; **29**: 703–712.
12. Kojima S, Nishihira K, Takegami M, Nakao YM, Honda S, Takahashi J, et al. Nationwide real-world database of 20,462 patients enrolled in the Japanese Acute Myocardial Infarction Registry (JAMIR): Impact of emergency coronary intervention in a super-aging population. *Int J Cardiol Heart Vasc* 2018; **20**: 1–6.
13. Yasuda S, Nakao K, Nishimura K, Miyamoto Y, Sumita Y, Shishido T, et al. The current status of cardiovascular medicine in Japan: Analysis of a large number of health records from a nationwide claim-based database, JROAD-DPC. *Circ J* 2016; **80**: 2327–2335.

Supplementary Files

Please find supplementary file(s);
<http://dx.doi.org/10.1253/circrep.CR-19-0106>