Integrative LHS for precision medicine research: A shared NIH and Taiwan CIMS experience

Ueng-Cheng Yang[1] | Tzu-Hung Hsiao[2] | Ching-Heng Lin[2] | Wei-Ju Lee[3,4,5,6] | Yu-Shan Lee[3,7] | Yang C. Fann[8]

[1] Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

[2] Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan

[3] Center for Geriatrics and Gerontology, Taichung Veterans General Hospital, Taichung, Taiwan

[4] Neurological Institute, Taichung Veterans General Hospital, Taichung, Taiwan

[5] Faculty of Medicine, National Yang-Ming University School of Medicine, Taipei, Taiwan

[6] Institute of Clinical Medicine, National Yang-Ming University School of Medicine, Taipei, Taiwan

[7] PhD Program in Translational Medicine, National Chung Hsing University, Taichung, Taiwan

[8] Division of Intramural Research, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland

Correspondence
Yang C. Fann, Division of Intramural Research, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA.
Email: fann@ninds.nih.gov

Abstract

Introduction: Precision medicine is an important milestone toward the attainment of personalized medicine. A learning health system (LHS) may facilitate the evidence collection and knowledge generation process for disease-based research and for the diagnosis, classification, or treatment of each disease subtype to improve patient care.

Methods: The LHS design and implementation used by Taichung Veterans General Hospital (TCVGH) in Taiwan for their newly funded precision medicine research, a dementia registry study, was modeled from an LHS developed at the National Institutes of Health in the United States. This Clinical Informatics and Management System (CIMS), including its subsystems, facilitates and enhances operations associated with the institutional review board, clinical research data collection and study management, the hospital biobank, and the participating health research centers to support their precision medicine research aimed at improving patient care.

Results: The implementation of a shared-design, full-cycle LHS with an enhanced CIMS, combined with hospital-based real-world data marts, has made the TCVGH dementia registry study a reality. The research data, including clinical assessment and genomics analysis information collected in CIMS, combined with data marts, are the foundation of the TCVGH dementia registry for outcome analyses. These high-quality datasets are useful for clinical validation, new hypotheses, and knowledge generation, leading to new clinical recommendations or guidelines for better patient treatment and care. The cyclic data flow supports the full-cycle LHS for TCVGH's dementia research to improve the care of elderly patients.

Conclusions: Knowledge generation requires high-quality research and health care datasets. While the details of LHS implementation methods in the United States and Taiwan may differ slightly, the LHS concept design and basic system architecture, with improved CIMSs, were proven feasible. As a result, learning health processes in support of translational research and the potential for improvement in patient care were significantly facilitated.

KEYWORDS

dementia, integrative information system, multiple sclerosis, precision medicine

1 | INTRODUCTION

Achieving the vision of personalized medicine through the precision medicine initiative[1] requires a comprehensive, multidisciplinary approach to support evidence-based biomedical research. Recent research advances in using big data combined with development of artificial intelligence (AI) have accelerated the pace of translational research. Therefore, the use of real-world data (RWD) such as diagnosis, treatment, and images in hospital information system (HIS), combined with insurance claim data, is becoming increasingly important to support evidenced-based translational research. While RWD are often not clean enough or optimized for direct research use, real-world "evidence"[2,3] is highly valued by major regulatory agencies around the world. Regulatory approval of diagnostic kits, devices, drugs, and biologics requires high-quality datasets from clinical trials to provide needed evidence. Furthermore, because the amount of data can be collected in a single institution is quite limited,[4] the formation of an interdisciplinary team and sharing of data are needed to obtain sufficient datasets to support rare diseases research[5] and precision medicine initiatives.[6]

Precision medicine research is a challenge not only due to data sharing requirements but also because certified genetic testing services that include diagnosis and interpretation are expensive. Other rate-limiting steps are the knowledge generation processes associated with diagnosis and classification of diseases to find optimal, individualized treatments and development of clinically validated biomarkers. It is, therefore, important to establish a full learning cycle ensuring that the learning is continuous, knowledge informs improvements, and a true learning health system (LHS) is formed.

Taichung Veterans General Hospital (TCVGH) has established five data marts from their de-identified Electronic Health Record (EHR) data warehouse: the geriatrics database, the cancer registry database (CRD), the health examination database, the health insurance claim database, and the internal clinical database (see Figure 1). These databases were established primarily for use by hospital administration and management and for investigator-initiated research. Any investigator may apply to receive de-identified data from these databases for use in research. For example, to propose and conduct any disease-specific translational study, such as a study of colon cancer, a new dataset can be easily constructed by using combinations of data from the five internal databases, upon Institutional Review Board (IRB) approval, without allowing investigators access to the entire data marts. Data from the CRD (similar to the Surveillance, Epidemiology, and End Results Program [SEER[7]] in the United States) includes the baseline information of cancer patients, such as characteristics, type, stage of cancer, and treatments. These data can be added to a new colon cancer database. For any descriptive study of cancers, the CRD is a great resource to begin to conduct a cross-sectional cohort study. Because this kind of dataset alone is not sufficient to achieve the goal of precision medicine,[8] the investigators often add high throughput data derived from biospecimens, thus, combining both genotype and phenotype information for further investigation. Note that the datasets derived from the HIS data marts combined with genomics data alone cannot really form a complete learning cycle without introducing a clinical study information system (CSIS) to facilitate study and patient management for the collection of high-quality research data. In this article, we describe our efforts to adopt an integrated LHS developed at the National Institute of Neurological Disorders and Stroke (NINDS) at National Institutes of Health (NIH). This system was modified for TCVGH to support their first dementia precision medicine research in Taiwan.

2 | RESEARCH INTERESTS

Principles and recommendations for sharing and reuse of "individual participant data" from clinical trials have been extensively discussed through a robust consensus process and endorsed by major institutions in the United States and European countries.[9] Taiwan is leading in the global open data index[10]; however, sharing of controlled data, such as health care data and genomic information, is not widespread. Therefore, the TCVGH has attempted to establish a de-identified, hospital-based registry project to support their new precision medicine research by implementing a LHS.[11]

TCVGH was established to care for an increasing number of aging veterans in Taiwan. This elderly population usually suffers from many age-related diseases that require medical attention and care from different clinical departments. To improve the quality of care for these veterans, in 2008, TCVGH established the Center for Geriatrics and Gerontology (CGG) to coordinate the resources of numerous clinical departments. The mission of the center is to provide patient-centered, holistic, interdisciplinary, and continuous geriatric care to elderly patients. Among age-related diseases, dementia is the most urgent
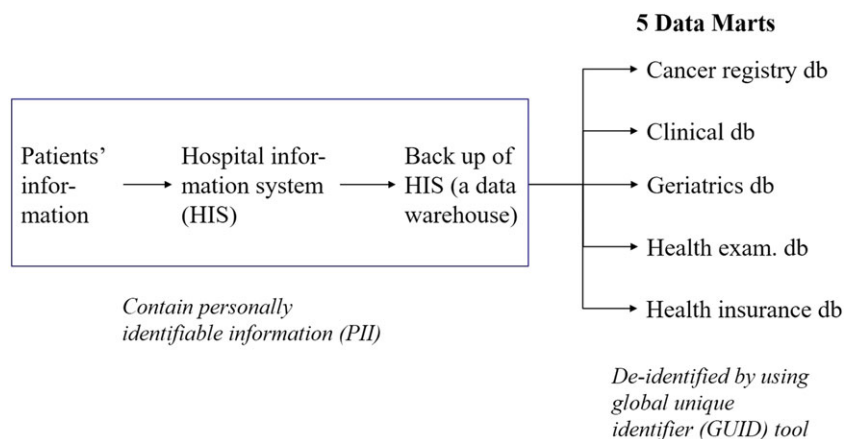

5 Data Marts

FIGURE 1 Existing clinical databases in Taichung Veterans General Hospital (TCVGH). Patients and health care information were entered into the hospital information system (HIS) by care physicians. A copy of the HIS was backed up to a data warehouse, which was then de-identified using a GUID tool in Clinical Informatics and Management System (CIMS). Information was then divided into five functional areas corresponding to hospital operations, which were then stored as separate data marts. These subsets are used by administration, hospital management, and research

and burdensome to veteran populations. It was thus a good starting point to use a precision medicine approach to establish a LHS to support translational research within this population.

The first step in establishing a LHS took advantage of existing information resources in the TCVGH in order to establish a hospital-based dementia registry. The new registry was designed not only to support new translational research but also to gain knowledge about diagnosis and treatment, in combination with lifestyle and e-health data, to improve patient care and facilitate data sharing for collaboration and physician education. This integrated registry is an important part of the learning cycle system, which provides the data for clinical validation of any proposed intervention studies and helps to discover new knowledge. Once research data are collected and analyzed and a new round of knowledge is discovered that influences and improves current patient care and future clinical practice, the next evidence-based RWD can then be collected and analyzed once again via the LHS cycle.

A second aim of this LHS was to provide the precision medicine infrastructure and services needed for future disease-based research. Thus, it is important to use all the available data—including the genetic and clinical information—to diagnose and treat the patients. Because biospecimens may bridge the clinical and genomic research studies, a hospital biobank was established and added to LHS as a key component in this learning process.

3 | METHODS

3.1 | The NIH CIMS experience

Achievement of the ultimate goals of precision medicine requires an integrative infrastructure with the ability to collect and validate diverse and high-quality individual research data, which is then shared among multidisciplinary teams to generate new knowledge and advance translational research aligned to the vision of LHSs. The major components of the integrative LHS used at NIH/NINDS, include (1)

the HIS (Clinical Research Information System [CRIS]), (2) a data warehouse (Biomedical Translational Research Information System [BTRIS][12]) and (3) a clinical research and study management system (Clinical Informatics and Management System [CIMS][13]). These are used for all NINDS clinical studies, including natural history research, clinical trials. In addition, there is a biobank and an integrated data portal with analytical tools.

The CIMS has four subsystems:

(1) a protocol tracking and management system (PTMS), which is an electronic institutional review board system (eIRB) developed to track and manage the life cycle of clinical research and trial protocols from submission to approval and for safety monitoring and reporting;

(2) a CSIS, which is an electronic data capture (EDC) and study management system used to facilitate the collection of high-quality clinical research data;

(3) a specimen tracking and management system (STAMS), which is used by the hospital biobank to manage and keep track of human subject specimens and their testing; and

(4) a Global Unique IDentifier (GUID[14]), which is used to de-link or de-identify personally identifiable information (PII), and which also allows the ability to associate the same patient across different studies. (Because of its unique features, the GUID can also be used for integrating de-identified datasets into a data repository.)

As an example, Figure 2 shows the full LHS cycle implementation of CIMS as used at NINDS for their intramural multiple sclerosis (MS) studies. Like most medical research institutions, patient care-related data were mainly collected in a HIS, in this case, the NIH Clinical Center's CRIS. To avoid any disruption of hospital operation and patient care, a separate data warehouse, the BTRIS, was created to provide clinical investigators with access to both identifiable data for subjects

FIGURE 2 A full LHS implementation of CIMS used at NIH/NINDS for intramural multiple sclerosis (MS) studies. This includes a hospital information system (CRIS), a data warehouse for translational research (BTRIS), and four subcomponents of the CIMS comprised of PTMS, CSIS, STAMS, and GUID as well as an integrated MS data portal with analytical tools for outcome validation, knowledge generation, and external data sharing/collaborations. BTRIS, Biomedical Translational Research Information System; CIMS, Clinical Informatics and Management System; CRIS, Clinical Research Information System; CSIS, clinical study information system; GUID, Global Unique IDentifier; LHS, learning health system; NIH, National Institutes of Health; NINDS, National Institute of Neurological Disorders and Stroke; PTMS, protocol tracking and management system; STAMS, specimen tracking and management system
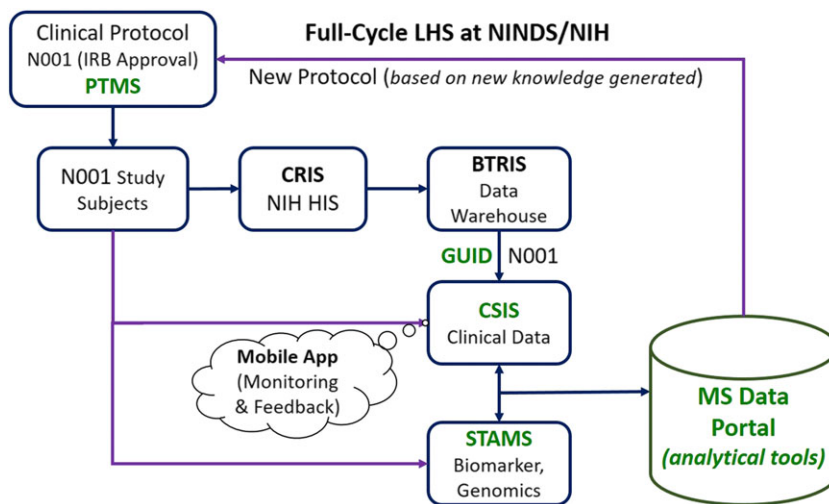
on their own active protocols and de-identified data for sharing among investigators for data analysis protocols. In addition to HIS data, all other protocol-related research data and biospecimens were collected in the CSIS and STAMS subsystems. The CSIS is also capable of collecting Internet of things (IoT)-based research data. In this MS study example, a mobile app designed to actively monitor and collect real-time patient movement and quality of life (QOL) feedback data sends this information directly to the investigators and research support staff for monitoring or for immediate attention, if required. Both biomarker and genomics data derived from the specimen can be combined with CSIS research data into an MS data portal for analysis and collaboration. The portal contains analytics tools, including R, statistics and machine learning algorithms to assist investigators in visualizing, correlating, and modeling their combined datasets for better treatment recommendations, and helps to generate new outcome-based knowledge to improve patient care and hypotheses for future clinical research. Thus, a continuous full LHS cycle is achieved. This system has been implemented and enhanced for over 10 years to support intramural MS studies and extramural collaborations with researchers in the United States and around the world.

3.2 | The TCVGH information system

The CIMS described in the previous section was obtained from NIH/NINDS through a material transfer agreement. TCVGH's LHS implementation was revised and enhanced to meet local needs in Taiwan, and it has been widely used by the National Research Program for Biopharmaceuticals, the largest clinical trial research network funded by Ministry of Science and Technology in Taiwan, to support all their funded clinical research studies.

Similar to NINDS's full-cycle LHS implementation, TCVGH already has an HIS and five data marts as described in Figure 1. To complete the LHS implementation to support TCVGH's vision for their precision medicine research in dementia, a modified CIMS infrastructure was added. For example, the PTMS review workflow was modified to adopt Taiwan's IRB regulations; the CSIS was modified to include Chinese case report form (CRF) and data validation against their local dementia protocol study design and rule sets; the STAMS was revised to allow their research support staff to collect multiple biospecimens in "cart" and then process it in batches for greater operational efficiency. Finally, the GUID was enhanced to allow usage of the National Health Insurance ID card to de-identify and generate GUID without manual data entry. (The NIH/NINDS implementation required this). The enhancements, as described below, have greatly increased the efficiency and adoption of the CIMS in Taiwan and implementation of the LHS in the TCVGH.

3.3 | The implementation and operation of LHS for the TCVGH dementia study

### 3.3.1 | Study management and data collection via CIMS

The multidepartment dementia study protocol was reviewed and received broad consensus after which it was submitted, reviewed, and approved by TCVGH IRB using the Taiwan PTMS system.

Qualified dementia patients were then registered in the CSIS and scheduled to visit different investigators who conducted different interventions and study plans at the TCVGH CGG (eg, Department of Psychiatry, Department of Neurology). The study CRFs were created in CSIS to collect standardized and consistent high-quality data, including patient information, past diseases, and medical history entered by different Principle Investigator (PI)s in the various departments. Data from the evaluation of neural and psychological functions, such as the Mini-Mental State Examination (MMSE) test,[15,16] Montreal Cognitive Assessment (MoCA),[17,18] Wisconsin Card Sorting Test (WCST[19]), was also readily collected.

To ensure the quality of data, the CSIS was revised to collect and validate the essential clinical information for qualified cohorts. As shown in Figure 3, routine clinical care data for dementia patients was collected from a de-identified research resource, called the "geriatrics database." This database is an internal patient data mart, which routinely collects the geriatrics study-related data from the main HIS in TCVGH. This geriatric database, which is essentially a hospital-based disease registry that collects data from patients with multiple neurological diseases, which also includes data from all dementia patients in this study. The geriatric database can be used as a research data mart after replacing the PII with a GUID, which was also used to integrate with research data, collected in CSIS. In addition to medical care and clinical research data collected, the mobile health data from the IoT (eg, wireless connected blood pressure monitors) may be collected and tracked in near real time in CSIS using web-based application programming interfaces (API). Some unstructured data, such as progress notes or radiology reports, can be processed through natural language processing (NLP) tools to be autocurated into CSIS. These capabilities greatly expand the ability of investigators and medical care staff to monitor and provide needed care for patients participating in research studies. The CSIS plays a key role in the TCVGH LHS implementation by facilitating not only the data capture process but also the quality assurance process.

With advanced sequencing technology, high throughput genomics data are now routinely generated at TCVGH laboratories using collected biospecimens in STAMS. In addition, the STAMS was enhanced to integrate clinical workflow from sample collection, bar-code generation and subsequent freezer storage, and routine lab sample processing into the study management pipeline. This greatly enhanced the clinical staff's efficiency and reduced errors during the clinical operations that supported this study. The specimen collection process was designed to use bar-code systems to ensure the sample data tracking quality. This bar-coded specimen identifier can be easily mapped to the GUID in a many-to-one relationship. This relationship can be relayed to any high throughput data generated from genomic studies conducted by bioinformatics PIs in the study. To make this research information available to researchers, the dementia registry portal was created to integrate CSIS and STAMS into a biobank information system to facilitate genotyping and phenotyping correlation analyses that are critical for precision medicine research. Both the processed and raw data of genomic studies were integrated in the dementia registry repository, which was used as an open and shared research resource. By linking the genotype (genomic data) with phenotype (clinical data), potential biomarkers for dementia and potential improved clinical care for study participants can be readily ascertained.

FIGURE 3 The TCVGH implementation of full cycle LHS for a dementia registry study was similar to that developed at NIH/NINDS. The dementia registry contains data from health care and lab tests, de-identified data from HIS in data marts, clinical research data (including mobile health data) collected through CSIS, and high throughput genomics and biomarker data derived from the biospecimens stored in STAMS. The integrated registry data combined with the analytical tools provide outcome validation and treatment recommendations, forming a feedback loop into the next LHS cycle. Note: Subsystem* indicates the individual customized and enhanced version from the NIH CIMS subsystem. CIMS, Clinical Informatics and Management System; CSIS, clinical study information system; GUID, Global Unique IDentifier; LHS, learning health system; NIH, National Institutes of Health; NINDS, National Institute of Neurological Disorders and Stroke; PTMS, protocol tracking and management system; STAMS, specimen tracking and management system; TCVGH, Taichung Veterans General Hospital
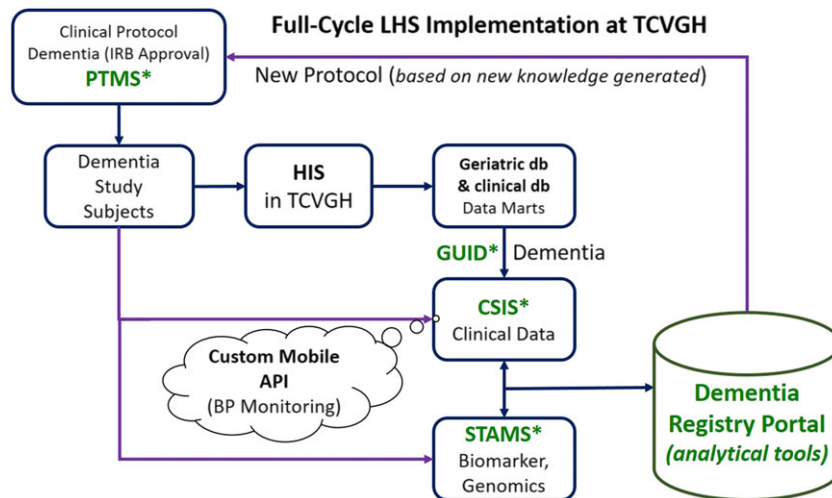
3.3.2 | Data standardization for data integration, sharing, and regulatory reporting

To reach the goal of data integration for sharing to accelerate precision medicine research, it is necessary to employ controlled vocabulary (such as the common data element [CDE][20]) or international data standards (such as Clinical Data Interchange Standard Consortium [CDISC] standards[21]). Most clinical software or tools were designed to facilitate data collection and quality through edit checks without optimization for data reuse or data mining and analysis. Clinical trial data for regulatory submissions in the United States and Japan are required to be standardized, based on CDISC standards. Standards-compliant data can be directly archived in a "repository" for data sharing and BIG data analytics. However, the effort required for after-the-fact data transformation is considerable and not easily attainable by investigator-initiated trials (IIT) wherein both funding and resources are limited. Therefore, it is important to implement and utilize standards before the start of a clinical trial.

The CIMS that supports the IIT in Taiwan was enhanced with a new "data standardization" function to make CDISC-compliant CRFs within the data collection module.[22] The study data can be collected in typical electronic data capture (EDC) form with the data elements being CDISC compliant. This new feature not only creates CDISC-compliant CRFs, which are useful in data interchange and sharing, but also in preparation of documents for regulatory submission. In addition, CSIS can import or export CRF data in an Operational Data Model (ODM) format,[23] which is a CDISC data transport standard. Many existing files can be reused or shared directly in a new research study. With a CDISC-compliant CSIS, TCVGH investigators can take the advantage of CDISC standards from the start of a clinical trial and tabulate trial data and files readily to produce the CDISC SDTM format[24] to facilitate their FDA submissions and expand international collaboration.

3.3.3 | Knowledge generation and LHS feedback process

Prior to new knowledge generation, which is a slow but essential process of any full-cycle of LHS, it is important for a clinical researcher to use collected research evidence to assess and improve patient care for better outcomes. The evidence collected in a hypothesis-driven research study or a clinical trial, such as the dementia registry study at TCVGH; the RWD, such as those recorded in the HIS or via IoT; the research data collected via standards such as CDISC-compliant CSIS CRFs; and high throughput data from the hospital laboratories must all be integrated and verified by a quality assurance or validation processes to support outcome assessment and knowledge discovery.

In a manner similar to that of the NIH/NINDS LHS implementations for neurological diseases, the datasets from TCVGH HIS data marts, CSIS databases, and biospecimen-generated data were integrated into a dementia registry using the GUID. With built-in analytical tools such as R and statistical algorithms, the data portal allows the researchers to visualize, assess, and analyze study outcomes interactively as well as to perform an individual patient's clinical assessment or validation based on genotype and phenotype datasets to generate a new personalized decision for medical treatments. If a potential clinical outcome is indeed recommended and changes the course of treatment for the patient (such as a change of guideline[25]), the new treatment evidences and results in routine medical service will then become RWD to support future clinical decisions. The continuous learning effort of this new clinical practice, treatment, and outcome cycles can be studied and followed up by another new clinical

research study, forming a knowledge feedback cycle to improve future medical care and achieving the goal of precision medicine research. With this LHS implementation, we were able to use an integrated data source for the analyses and knowledge generation processes (D to K). The newly generated knowledge provides optimized guidance for clinical practice (K to P). The data generated from the new clinical practice will again go into the dementia registry (P to D) for the next round of the learning health cycle. The integrated portal for data sharing and analysis of multiple data sources enabled "team science" research and greatly enhanced efficiency for collaboration in the dementia registry study at TCVGH toward achieving the goals of the LHS.

3.4 | Implications and lessons learned

A medical research center such as TCVGH has three primary objectives: medical practice, patient care, and research. The patient records in HIS and the health data collected by the IoT as part of a research study were considered to be RWD. RWD are useful but often, the quality is not satisfactory or not sufficient for clinical research. The success of this large undertaking involved an effort to initiate precision medical research in a hospital care environment, using biospecimens and genomic data from consenting patients and integrating this information with clinical research data. The existing data marts at TCVGH were not sufficient to support the diverse data sources, data integration, and improvement in the quality of RWD and its continuous analyses for knowledge discovery; however, combining this with high-quality research and the adaptation of a well-developed LHS implementation from the NIH/NINDS has been quite successful.

The full cycle of the LHS implementation at TCVGH is similar to that at NIH/NINDS, taking advantage of the experience and infrastructure sharing between two countries and institutions. The lessons learned and the knowledge shared from each of these two implementations, along with infrastructure improvements, greatly enhanced and accelerated biomedical research in both countries in the area of precision medicine.

The key success of the TCVGH LHS implementation largely relied on the roles of CIMS with its subsystems in facilitating the protocol approval, standard data collection, biospecimen management, quality assurance, and data integration. TCVGH built the dementia registry and portal for data sharing among different investigators and supported each step of the LHS process. The enhanced CIMS was able to effectively collect and link different types of information through standardized CDISC-compliant eCRFs, and GUID was used to de-identify and integrate datasets. Data quality was improved by using the data validation and data standard control functions in CIMS; thus, datasets could be readily integrated for the analytical pipeline and knowledge discovery processes. The clinical operation and support burdens with biospecimen collection and sample processing during the patient visits were greatly reduced by seamlessly linking the biospecimen information stored in STAMS with the associated clinical information collected in CSIS via the GUID. While the current knowledge discovery process depends largely on the interactive and statistical tools embedded in the LHS cycles (mainly in the registry portal), efforts are underway to advance this process through AI tools. The

dementia registry study is only a beginning of the TCVGH efforts to initiate precision medicine research. Leveraging the full-cycle LHS implementation experience shared by NIH/NINDS helped TCVGH rapidly initiate this LHS implementation.

4 | CONCLUSIONS

The MS project and dementia registry are precision medicine research projects implemented by NIH and TCVGH, respectively. They have both demonstrated the feasibility of establishing a similar LHS implementation in a hospital research environment. Our successful strategy and experience can be readily replicated and customized to other projects in Taiwan, the United States, or around the world. Customizable learning cycles with CIMS can be further improved by inviting more interdisciplinary research teams or centers to join. For example, adding patients' reported outcomes research could be valuable in advancing personalized care. In addition, the LHS infrastructure could be easily expanded beyond TCVGH to include other veterans' care centers. Once the LHS culture is established, patient care will gradually and continuously improve over time.

One potential concern of a project-based LHS is its sustainability. For this reason, our LHS implementation was built on TCVGH's new precision medicine research, which maintains that an LHS is an efficient way to improve diagnosis and treatment through the use of RWD. When diagnosis and treatment become more precise, more patients will be attracted to TCVGH; this, in turn, will create more revenue to make the LHS sustainable. Currently, TCVGH is actively identifying other top diseases (such as diabetes and chronic kidney diseases) to include in new LHS implementations. This can help to ensure the sustainability of project-based LHSs.

CONFLICT OF INTEREST

All authors have affirmed that they do not have any conflicts of interest to declare.

ORCID

Yang C. Fann http://orcid.org/0000-0003-2636-2002

REFERENCES

1. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine*. Washington, D.C: National Academies Press; 2011. https://doi.org/10.17226/13284.

2. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther*. Wiley-Blackwell. 2018;103(2):202-205. https://doi.org/10.1002/cpt.946

3. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med*. 2016;375(23):2293-2297. https://doi.org/10.1056/NEJMsb1609216

4. Carter RE, Sonne SC, Brady KT. Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study. *BMC Med Res Methodol*. BioMed Central. 2005;5(1):11. https://doi.org/10.1186/1471-2288-5-11

5. Brownstein CA, Holm IA, Ramoni R, Goldstein DB, Members of the Undiagnosed Diseases Network. Data sharing in the undiagnosed

diseases network. *Hum Mutat*. 2015;36(10):985-988. https://doi.org/10.1002/humu.22840

6. Scollen S, Page A, Wilson J, on behalf of the Global Alliance for Genomics and Health. From the data on many, precision medicine for "one": the case for widespread genomic data sharing. *Biomedicine Hub*. Karger Publishers. 2017;2(1):21-21. https://doi.org/10.1159/000481682

7. Altekruse SF, Rosenfeld GE, Carrick DM, et al. SEER cancer registry biospecimen research: yesterday and tomorrow. *Cancer Epidemiol Biomarkers Prev*. 2014;23(12):2681-2687. https://doi.org/10.1158/1055-9965.EPI-14-0490

8. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med*. 2016;8(1):134. https://doi.org/10.1186/s13073-016-0388-7

9. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. British Medical Journal Publishing Group. 2017;7(12): e018647. https://doi.org/10.1136/bmjopen-2017-018647

10. Open knowledge network 2017. *Place overview—Global Open Data Index*. Available at: https://index.okfn.org/place/ (Accessed: 4May2018).

11. Friedman C, Rubin J, Brown J, et al. Toward a science of learning systems: a research agenda for the high-functioning learning health system. *J Am Med Inform Assoc: JAMIA*. 2015;22(1):43-50. https://doi.org/10.1136/amiajnl-2014-002977

12. Cimino JJ, Ayres EJ, Remennik L, et al. The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date. *J Biomed Inform*. 2014;52:11-27. https://doi.org/10.1016/j.jbi.2013.11.004

13. Cheung, H. Fann Y, Wang SA, et al. 'A web-based protocol tracking management system for clinical research', *Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems*, (may), 2004; 72–77. https://doi.org/10.1109/CBMS.2004.1311694.

14. Johnson SB, Whitney G, McAuliffe M, et al. Using global unique identifiers to link autism collections. *J Am Med Inform Assoc*. 2010;17(6):689-695. https://doi.org/10.1136/jamia.2009.002063

15. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. *J Am Geriatr Soc*. 1992;40(9):922-935. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1512391. Accessed: 4May2018

16. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1202204. Accessed: 4May2018

17. Nasreddine ZS, Phillips NA, Bédirian Vé, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

18. Tsai C-F, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. *Int Psychogeriatr*. 2012;24(04):651-658. https://doi.org/10.1017/S1041610211002298

19. Stuss DT, Levine B, Alexander MP, et al. Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*. 2000;38(4):388-402. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10683390 (Accessed: 4May2018)

20. Sheehan J, Hirschfeld S, Foster E, et al. Improving the value of clinical research through the use of common data elements. *Clin Trials: J Soc Clin Trials*. 2016;13(6):671-676. https://doi.org/10.1177/1740774516653238

21. Souza T, Kush R, Evans JP. Global clinical data interchange standards are here! *Drug Discov Today*. 2007;12(3–4):174-181. https://doi.org/10.1016/j.drudis.2006.12.012

22. Lin C-H, Chou H-I, Yang U-C. A standard-driven approach for electronic submission to pharmaceutical regulatory authorities. *J Biomed Inform*. 2018;79:60-70. https://doi.org/10.1016/j.jbi.2018.01.006

23. Huser V, Sastry C, Breymaier M, Idriss A, Cimino JJ. Standardizing data exchange for clinical research protocols and case report forms: an assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). *J Biomed Inform*. 2015;57:88-99. https://doi.org/10.1016/j.jbi.2015.06.023

24. Wood F, Guinter T. Evolution and implementation of the CDISC Study Data Tabulation Model (SDTM). *Pharma Program*. 2008;1(1):20-27. https://doi.org/10.1179/175709208X334623

25. Liou J-M, Chen CC, Chen MJ, et al. Sequential versus triple therapy for the first-line treatment of *Helicobacter pylori*: a multicentre, open-label, randomised trial. *Lancet*. 2013;381(9862):205-213. https://doi.org/10.1016/S0140-6736(12)61579-7