MJ Health Research Foundation

MJ Health Resource Center

The introduction of
MJ Health Database

Technical Report No. : MJHRF-TR-01

2016/08/22

The introduction of MJ Health Database

I.   Introduction

In recent years, significant global disease burden is found to be attributed to the non-communicable diseases (NCD). According to the WHO Global Status Report on non-communicable diseases 2014, NCDs cause more deaths than all other deaths combined and are projected to cause 52 million deaths by 2030 (1). Some of these non-communicable diseases are multifactorial and can develop as a result of both genetic predispositions and environmental factors (2, 3). In order to better understand the disease etiology and disease progression mechanism, large scale health datasets with accompanying biological samples were collected to estimate potential associations between physiological, socioeconomic or genetic factors and chronic diseases (4, 5). MJ Health Database (MJHD) is a longitudinal, population-based health research database consisting of socioeconomic, behavioral and physical examination data. As will be clear in the following section, the volume and diversity of data stored in the MJHD, as well as the associated biological samples essentially satisfy material requirements for most large-scale prospective studies.

II.   MJ Health Database

MJ Health Database (MJHD) is a longitudinal, large scale, comprehensive, population-based health data. Participants of MJHD were apparently healthy individuals who received physical examination services at MJ Health Management Institution, a private healthcare firm in Taiwan, since 1994. In addition to the health data, some participants agreed to donate blood samples as early as 2002. MJ Health Management Institution has four Health Screening Centers in Taiwan, located in Taipei (northern region), Taoyuan (northwestern region), Taichung (central region), and Kaohsiung (southern region), providing comprehensive physical examination service to the general public in Taiwan and neighboring region in Asia. In terms of longitudinal breadth, MJHD contains the results of actual, ongoing physical examination service dating back to 1994 and has no pre-defined termination date. Many participants had multiple visits over the years. For each examination visit, information on behavior and lifestyle was collected via health questionnaire, while data on anthropometric and biological tests were collected via physical examination. The earliest examination recorded in MJHD was in 1994, while the earliest health questionnaire in 1996. Between 1996 and the end of 2015, MJHD has accumulated over 1.5 million

examination and questionnaire records for 665,156 individual participants, with those aged 20-and-above constituting over 3% of the 20-and-above adult population in Taiwan (6). Due to the fact that people can participate in physical examination at the MJ Health Screening Centers either via family based membership or through employer sponsorship, MJHD participants can be grouped into either family or firm clusters depending on their participation mechanism.

The current MJHD participants reflect cultural and ethnic diversities found in both China and Taiwan. Taiwan is a multicultural society, with immigrants coming from diverse parts of mainland China at different points in time since the 17th century, in addition to the indigenous Malayo-Polynesian peoples who have lived on the island for millennia (7). The convergence of Chinese, aboriginal, Dutch, Spanish and Japanese cultures in Taiwan also makes it an ideal laboratory for research on socio-cultural changes in Chinese society (8). To accentuate Taiwan's unique population characteristic and to aid data users on utilizing this information, MJHD keeps records of ethnic backgrounds not only for all the participants, but also for their parents and grandparents, via self-reported questionnaire.

III. Health Survey – socioeconomic, demographic and lifestyle/behavioral data

Each physical examination participant completed a self-administered questionnaire prior to his or her examination appointment. The questionnaire has undergone several revisions over the years, and currently consists of about 130 multiple choice questions, which are divided into 7 areas: basic information, personal and family medical history, current health status, lifestyle, physical exercise, sleep habits, and dietary habits (Table 1). Data collected through questionnaires were stored as electronic record since 1996.

Table 1. MJHD socioeconomic, demographic and lifestyle/behavioral data summary

Question Type	Detail
Basic information	Age, occupation, education, family income, race/ethnicity
Personal and family medical history	Cancer, chronic diseases
Current health status	Tinnitus, vertigo, bowel movement, digestive system disorders, etc.
Lifestyle/behavior	Smoking, drinking, betel nut chewing, etc.
Physical exercise	Frequency and duration
Sleep Habits	Quality and duration

Dietary Habits	Food/nutrition intake, frequency and amount

IV. Biodata – physical examination data

Physical examinations performed at MJ Health Screening Centers covered over 100 important biological indicators, including anthropometric measurements, blood tests, imaging tests, etc. (Table 2). Venipuncture was performed to collect overnight fasting blood for a series of biochemical analyses, including lipid profile, renal function, liver profile, serum albumin, globulin, and uric acid analyses, etc. Fecal occult blood test and urine analysis were also performed. All biological specimens were analyzed in a laboratory located in each MJ Health Screening Center. Other physical examination tests such as abdominal ultrasounds, X-rays, electrocardiogram, lung function, hearing, vision, intraocular pressure, retinal imaging, gynecological examination, and bone mineral density were also performed. A standardized service workflow was adopted by all four Health Screening Centers in Taiwan, which conformed to ISO 9001-2008 requirements for quality management systems. MJ Health Screening Center laboratories have participated in the College of American Pathologists (CAP) Proficiency Testing to ensure their ability to generate quality health data. As laboratory equipment can change over time, detailed records of equipment manufacturers, model numbers and period of use were kept in the technical report for Laboratory Equipment, MJHRF-TR-06 (9).

Table 2. MJHD biodata summary

Data Type	Detail
Anthropometric measurements	Height, weight, body fat, waist circumference, hip circumference, blood pressure, pulse, etc.
Blood tests	Complete blood count, blood glucose, blood lipid, liver function, kidney function, urea, thyroid function, calcium/phosphorus/iron, hepatitis B/C, inflammation factors, tumor markers, etc.
Stool and urine analyses	Protein, glucose, fecal occult blood test, etc.
Ultrasounds	Breasts, abdominal, etc.
X-rays	Chest, abdomen, upper gastrointestinal series
Other	Electrocardiogram, lung function, vision tests, gynecological examination, bone mineral density, etc.

V. MJ Biobank

In addition to questionnaire and physical examination data, MJHD also has an accompanying biobank, providing the biomedical research community a new avenue for population-based longitudinal research on the molecular level. The biobank stores blood samples donated by physical examination participants. Sample collection first started in 2002, and can be divided into two phases: 2002 to 2008 and 2015 to present. As of the end of 2015, MJ Biobank has collected 558,104 serum, 221,397 plasma and 261,861 buffy coat samples from 84,542 unique individuals. Blood samples were stored in -80°C freezers on a long term basis. As of the end of 2015, 1,834 have self-reported having developed cancer in the past (Table 3).

Table 3. Distribution of self-reported cancer at least once among blood sample donors as of end of 2015

Cancer Type	Male	Female	Total
Nasopharyngeal Cancer	48	27	75
Lung Cancer	38	39	77
Breast Cancer	1	264	265
Stomach Cancer	39	18	57
Liver Cancer	81	24	105
Rectal Cancer	116	63	179
Cervical Cancer	0	530	530
Prostate Cancer	131	0	131
Other Cancers	254	271	525
Total	666	1,168	1,834

VI. Major characteristics of data

MJHD health data and biological samples were voluntarily donated by physical examination participants. Therefore, the total number of donations, the frequency of each donation, as well as the time interval between each donation varies from one individual to another. Among a total of 665,156 physical examination participants, 381,032 or 57.28% had only one visit, while others had at least 2 visits or more. Based on the length of time between the first and most recent visits for any given participant, among the 284,124 participants who had at least 2 visits, approximately 52% had stayed 4 years or less with the cohort, while 22% stayed 10 years or longer. The distribution of select demographic and social characteristics of participants as recorded at the first physical examination is shown in Table 4.

Table 4. Demographic Distributions of MJHD based on each participant's first
examination visit

	Male (n=318,356)		Female (n=346,800)		Total (n=665,156)	
N of Visits(1-31)	2.3	(2.5)	2.2	(2.3)	2.3	(2.4)
Age(1-98)	39.3	(14.4)	39.5	(14.3)	39.4	(14.4)
0-19	15,123	4.8%	12,966	3.7%	28,089	4.2%
20-64	281,430	88.4%	314,134	90.6%	595,564	89.5%
65 or older	21,803	6.8%	19,700	5.7%	41,503	6.2%
Total	318,356	100.0%	346,800	100.0%	665,156	100.0%
Education						
Junior High & Below	45,932	18.4%	74,049	27.5%	119,981	23.1%
Senior High	52,810	21.2%	62,215	23.1%	115,025	22.2%
College & Above	150,763	60.4%	133,181	49.4%	283,944	54.7%
Total	249,505	100.0%	269,445	100.0%	518,950	100.0%
Marriage						
No	90,376	37.0%	109,248	41.5%	199,624	39.3%
Yes	154,084	63.0%	153,964	58.5%	308,048	60.7%
Total	244,460	100.0%	263,212	100.0%	507,672	100.0%
Employment						
No	28,203	11.6%	93,542	35.9%	121,745	24.1%
Yes	215,215	88.4%	167,314	64.1%	382,529	75.9%
Total	243,418	100.0%	260,856	100.0%	504,274	100.0%

Standard deviations are in the parentheses.

In terms of blood sample collection, some MJHD participants donated his/her blood once, while others made multiple donations over the course of seven years. Because samples donated during any given examination visit can be linked to the donor's data profiles before and after the sample-donating visit, the dataset generated as a whole is thus longitudinal by nature. As shown in Figure 1, among those who donated samples between 2002 and 2008, 50% had physical examination visits prior to the visit when they donated samples, while 57% had visits after the sample-donating visits up until end of 2015. Average number of physical-examination-only visits were 2.8 and 3 per donor before and after sample-donating examination visit, respectively.
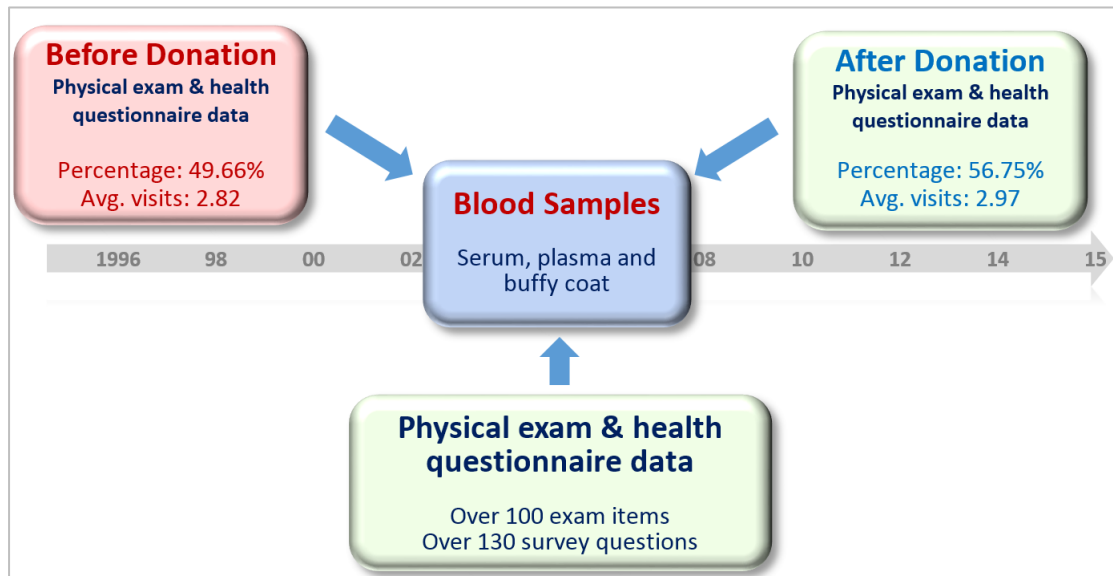
Fig 1. Percent donors who came for physical examinations before and after blood-donating visits

VII. Privacy Protection and Quality Control

All the personally identifiable information were removed from the raw data and replaced by a unique identification code. Personally identifiable information was then stored in a dedicated server and can only be accessed by authorized personnel. Details for data de-identification measures taken by MJHRF can be found in the MJHD technical reports for Information Security, MJHRF-TR-03 (10). To ensure the quality of the data released for public use, all data have gone through extensive data cleaning including logic check and illegal value treatment. For data user's convenience, the data cleaning process also included separating variables to the same question that have been harmonized across different questionnaire versions, etc. Full description for data cleaning measures are located in the MJHD technical reports for Data Cleaning, MJHRF-TR-04 (11).

VIII. Informed Consent and Re-consent for MJ Biobank

Explicit consents were obtained from the MJHD participants prior to the incorporation of their data and/or biological samples into the MJHD. Specifically, health questionnaires together with informed consent form were sent to all of the physical examination participants before their clinical appointment. The participants could choose to indicate whether they agree to transfer their health data to MJHRF for research purposes. Only data carrying proper consents would be included into the

MJHD.

Prior to 2010, laws or regulations governing the establishment and operation of biorepositories were lacking in Taiwan. In 2002, recognizing that no biobanking laws existed, the MJ Biobank obtained Institutional Review Board's approval and informed consents in the form of clinical trial and conducted its first phase of biological sample collection. The first phase of sample collection ended in 2008 when the government began to draft a biorepository law. In 2010, Taiwanese government enacted the Human Biobank Management Act or HBMA (12). HBMA is the legal framework that regulates the establishment, management and application of biobanks located in Taiwan. To comply with the new laws, an independent Ethics and Governance Council (EGC) was established to oversee all operations associated with the use of materials stored in the biobank. In September 2014, MJ Biobank obtained its official approval to operate from the Ministry of Health and Welfare. In light of its official approval, in October 2014, existing sample donors who donated samples between 2002 and 2008 were requested to re-consent by signing new informed consents. In the meantime, second phase of blood sample collection was initiated in January 2015. As of the end of July 2016, 11,252 re-consents were received, representing 15% of total re-consents.

IX.   Management of MJHD

MJ Health Research Foundation (MJHRF hereafter) is a non-profit health research organization in Taiwan, established in 2012 with funds donated by the MJ Group, the parent company of the MJ Health Management Institution. The mission of MJHRF is to manage the aforementioned health data and to promote cutting edge innovation in biomedical research. The ethnic-Chinese-specific health data might be of value to biomedical research in both academia community and industry setting. MJ Health Resource Center is an operational unit under MJHRF and is responsible for the storage and management of large-scale longitudinal health data derived from the comprehensive health examination services. MJHRF has a full range of security measures, advanced sample storage facility, and a set of stringent procedures with respect to the management and release of sensitive health data and biological samples. All of MJHRF's operations pertaining to the acquisition and application of its health resources are being governed by its Board of Directors. The full description of MJ Health Research Foundation and MJ Health Resource Center can be found at Website: http://www.mjhrf.org/main/page/about/en/#about01.

X.   Current Users and Publications

MJ Health Management Institution has been actively collaborating with academic and research institutions on various public health research projects since the 1990s (13). Between 2001 and 2016, utilizations of data from MJHD have led to 47 scholarly publications, some of which were in high impact journals. For instance, hepatocellular carcinoma risk prediction models were developed based on a cohort of 428,584, with age, sex, health history, hepatitis B and C viral infections, alanine transaminase (ALT), aspartate transaminase (AST), alfa-fetoprotein (AFP), etc. as model inputs (14). Furthermore, analyzing a cohort of 416,175 subjects, researchers reported the health benefit of leisure-time physical activity (LTPA) in the forms of reduced mortality and extended life expectancy. LTPA level for the study subjects were ascertained based on 3 questions from the MJHD health questionnaire: the types and intensities of weekly LTPA done during the previous month, duration of LTPA per week within the previous month, and the amount of physical activity done at work (15). In another study, MJHD data was incorporated as a part of a meta-analysis of 1.7 million subjects to study changes in estimated glomerular filtration rate (eGFR) and how it relates to end-stage renal disease (ESRD) and mortality. Age, sex, total cholesterol, systolic blood pressure, diabetes, history of cardiovascular disease, albuminuria and smoking status were included in the study (16). Those who are interested can visit http://www.mjhrf.org/main/page/resource/en/#resource07 for a list of publications.

XI.   Data resource access

MJHD data and biological samples are available for use by the research community worldwide. Health questionnaire and physical examination data are released on an annual basis, with the latest versions being made ready through the end of 2015. Codebooks with a list of variables covered in the MJHD data are available for download at http://www.mjhrf.org/main/page/resource/en/#resource03. To access MJHD data, researchers need to go through the application process and submit relevant documents as detailed at http://www.mjhrf.org/main/page/release2/en/#release01. For those interested in obtaining biological samples in addition to the data, details regarding biological sample application can be accessed at http://www.mjhrf.org/main/page/release1/en/#release01. Management or service fees will apply, and an official authorization code will be issued when MJHD resource is released. Any inquiries regarding MJHD data and biological sample access policies and application guidelines can be directed to the following address: contact_us@mjhrf.org.

References

1). World Health Organization. Global Status Report on noncommunicable diseases 2014. World Health Organization. 2014; Available: http://www.who.int/nmh/publications/ncd-status-report-2014/en/.

2). World Health Organization. Genes and Human Disease – Genes and noncommunicable diseases. World Health Organization. Available: http://www.who.int/genomics/public/geneticdiseases/en/index3.html.

3). Sing CF, Stengård JH, Kardia, SLR. Genes, Environment, and Cardiovascular Disease. Arterioscler Thromb Vasc Biol. 2003; 23: 1190-96.

4). Chen Z, Chen J, Collins R, Guo Y, Peto R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011; 40: 1652-66.

5). Jiang C, Thomas GN, Lam TH, Schooling CM, Zhang W, et al. Cohort Profile: The Guangzhou Biobank Cohort Study, a Guangzhou-Hong Kong-Birmingham collaboration. Int J Epidemiol. 2006; 35: 844-52.

6). Department of Household Registration, M.O.I. Household registration statistics data analysis in December 2015. Ministry of the Interior. Available: http://www.ris.gov.tw/en/web/ris3-english/home/-/asset_publisher/Ebr8/content/household-registration-statistics-data-analysis-in-december-2015?redirect=http%3A%2F%2Fwww.ris.gov.tw%2Fen%2Fweb%2Fris3-english%2Fhome%3Fp_p_id%3D101_INSTANCE_Ebr8%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_pos%3D1%26p_p_col_count%3D2.

7). Department of Information Services, Executive Yuan. The Republic of China Yearbook 2015 Taiwan. Executive Yuan. Available: http://www.ey.gov.tw/en/cp.aspx?n=6CF4694061D32E34.

8). Chen SH. Taiwan as a laboratory for the study of Chinese society and culture. Bulletin of the Institute of Ethnology (Academia Sinica). 1966; 14: 1-14.

9). Wang, ML. MJ Health Research Foundation Technical Report – Laboratory Equipment (MJHRF-TR-06). MJ Health Research Foundation. 2016; Available: http://www.mjhrf.org/main/page/resource/en/#resource08.

10). Chuang, YC. MJ Health Research Foundation Technical Report – Information Security (MJHRF-TR-03). MJ Health Research Foundation. 2016; Available: http://www.mjhrf.org/main/page/resource/en/#resource08.

11). Chuang, YC. MJ Health Research Foundation Technical Report – Data Cleaning (MJHRF-TR-04). MJ Health Research Foundation. 2016; Available: http://www.mjhrf.org/main/page/resource/en/#resource08.

12). Ministry of Health and Welfare. Human Biobank Management Act. Laws & Regulations Database of the Republic of China. 2012; Available: http://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020164.

13). MJ AMHTS. The Annual Report of MJ Health Screening Center 2012. MJ Health Management Institution; 2013.

14). Wen CP, Lin J, Yang YC, Tsai MK, Tsao CK, Etzel C, Huang M, Hsu CY, Ye Y, Lopa M, Hawk E, Wu X. Hepatocellular carcinoma risk prediction model for the general population: the predictive power of transaminases. J Natl Cancer Inst. 2012; 104: 1599-1611.

15). Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet. 2011; 378: 1244-53.

16). Coresh J, Turin TC, Matsushita K, Sang Y, Ballew SH, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. JAMA. 2014; 311: 2518-31.