Data Sources for Precision Medicine Research in Korea

Hyeon Chang Kim, MD, PhD
Department of Preventive Medicine
Yonsei University College of Medicine, Seoul, Korea
CONTENTS

• Precision Medicine (PM), PMI & PMI cohort
• Conventional cohort & PMI cohort
• Healthcare Big Data for Precision Medicine in Korea
PM, PMI, and PMI cohort?

**Precision medicine (PM)** is a medical model that proposes the customization of healthcare, with medical decisions, practices, and/or products being tailored to the individual patient.

It does not mean the creation of drugs or devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease, in the biology and/or prognosis of those diseases they may develop, or in their response to a specific treatment.

Tools employed in precision medicine can include genetic/molecular diagnostics, imaging, and analytics/software.
PM, PMI, and PMI cohort?

Precision Medicine Initiative (PMI) is a research project created by Barack Obama in 2015 with $215 million in funding that aims to make advances in tailoring medical care to the individual.

PMI Cohort Program is a landmark longitudinal research effort that aims to engage 1 million or more U.S. participants to improve our ability to prevent and treat disease based on individual differences in lifestyle, environment and genetics.
THE PRECISION MEDICINE INITIATIVE® (PMI)

Announced by President Barack Obama in his 2015 State of the Union address

MISSION: To enable a new era of medicine through research, technology, and policies that empower patients, researchers, and providers to work together toward development of individualized care.
About the Precision Medicine Initiative Cohort Program

Far too many diseases do not have a proven means of prevention or effective treatments. We must gain better insights into the biological, environmental, and behavioral influences on these diseases to make a difference for the millions of Americans who suffer from them. Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.
ALL OF US RESEARCH PROGRAM

All of Us Research Program

About the All of Us Research Program

Far too many diseases do not have proven preventions or treatments. To make a difference for the millions of Americans who suffer from them, we must gain better insights into the biological, environmental, and behavioral factors that drive these diseases. Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in environment, lifestyle and genes for each person.

As part of the federal Precision Medicine Initiative, the NIH is leading the effort to build a national, large-scale research enterprise with one million or more volunteers to extend precision medicine to all diseases. The All of Us Research Program, formerly known as the...
A TRANSFORMATIONAL APPROACH TO PARTICIPATION

Participants in the PMI Cohort Program will be true partners—not patients, not subjects—in the research process.

Involved in every step of program development:

- What data we collect
- What lab analyses we do
- What research is conducted
- How data gets returned
A TRANSFORMATIONAL APPROACH TO DIVERSITY

The cohort will reflect the rich diversity of America to produce meaningful health outcomes for historically underrepresented communities

- All economic levels
- All races and ethnicities
A TRANSFORMATIONAL APPROACH TO DATA

• **Data sharing will be swift** to both researchers and participants

  • Participants will have access to study information and data about themselves

• Data collection will start small and will grow over time

• Privacy and security will adhere to the highest standards
The Program will start by collecting a limited set of standardized data from sources that will include:

- Participant questionnaires
- Electronic health records
- A baseline physical evaluation
- Biospecimens (blood and urine samples)
- Mobile/wearable technologies
- Geospatial/environmental data

Data types will grow and evolve with science, technology, and trust.
INFORMATION FLOW IN

Direct Volunteers

- Self-report Measures
- mHealth Data
- Consent
- EHR Data
- Baseline Exam
- Biospecimens

Cloud Storage

HPO Volunteers
Volunteers

Data
- Individual data
- Individual health information
- Study updates
- Aggregated results

Public

Results

INFORMATION FLOW OUT
BENEFITS OF APPROACH

• Large and diverse
  • Less costly and less difficult than representative sample (which is rarely achievable)
  • Able to generate estimates of effect/association
  • Permits well-powered studies

• Support focus on underserved and underrepresented populations

• Complement (not duplicate) existing disease-specific cohorts
CONTENTS

• Precision Medicine (PM), PMI & PMI cohort
• Conventional cohort & PMI cohort
• Healthcare Big Data for Precision Medicine in Korea
Cohort studies are still needed in the era of PM.

Cohort studies are best for investigating....

- Incidence, prevalence, and their trends.
- Identifying population-specific issues.
- Finding new risk factors, biomarkers, and predictors in populations or patients.
- Rare adverse effects & pleiotropic effects of drugs.
- Combined effects of multiple drugs and non-drug treatments.
Cohort studies are still needed in the era of PM.

Cohort studies have advantages...

• Less ethical problems
• Not enough data to design an RCT
• Many research hypotheses can be tested

Cohort studies are strengthened with...

• Bio-banks
• Hybrid designs
• Multi-center cohorts, Multi-cohort analysis
## Randomized trial vs. cohort studies

<table>
<thead>
<tr>
<th></th>
<th>Randomized trial</th>
<th>Cohort study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In theory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment (exposure)</td>
<td>Random</td>
<td>Non-random</td>
</tr>
<tr>
<td>Patients (participants)</td>
<td>Non-random</td>
<td>Random (possibly)</td>
</tr>
<tr>
<td><strong>In actual design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypotheses</td>
<td>One (or a few)</td>
<td>Several</td>
</tr>
<tr>
<td>Participants</td>
<td>Homogeneous</td>
<td>Somewhat heterogeneous</td>
</tr>
<tr>
<td>Typical size</td>
<td>Hundreds to thousands</td>
<td>Thousands or more</td>
</tr>
<tr>
<td>Baseline data</td>
<td>Limited to essential data Collected usually once</td>
<td>Various but pre-determined Can be expanded later Collected regularly</td>
</tr>
<tr>
<td>Follow-up data</td>
<td>Outcomes of interests, Deaths, Pleiotropic effects</td>
<td>Selected disorders, Intermediate outcomes, Deaths, ...</td>
</tr>
</tbody>
</table>
Cohort studies are producing bigger data

- Cohort sample size is growing
- Expanding data types
- Genome, metabolome, microbiome, ...
- Digitalized imaging data
- Information & communication technologies
## Conventional Cohort vs. Precision Medicine Cohort

<table>
<thead>
<tr>
<th></th>
<th>Conventional cohort</th>
<th>PMI cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In theory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposures</td>
<td>Selected</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Participants</td>
<td>Random (possibly)</td>
<td>Non-random, but diverse</td>
</tr>
<tr>
<td><strong>In actual design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypotheses</td>
<td>Several</td>
<td>Not pre-determined</td>
</tr>
<tr>
<td>Participants</td>
<td>Somewhat heterogeneous</td>
<td>Seeking diversity</td>
</tr>
<tr>
<td>Typical size</td>
<td>Thousands or more</td>
<td>Hundred thousands to millions</td>
</tr>
<tr>
<td>Baseline data</td>
<td>Pre-determined &amp; Later-expanded</td>
<td>Unlimited &amp; Increasing</td>
</tr>
<tr>
<td></td>
<td>Collected regularly</td>
<td>Collected continuously</td>
</tr>
<tr>
<td>Follow-up data</td>
<td>Selected disorders, Intermediate</td>
<td>All available information</td>
</tr>
<tr>
<td></td>
<td>outcomes, Deaths, ...</td>
<td></td>
</tr>
</tbody>
</table>
Precision Medicine Cohorts needs Big Data Analytics

- **Volume**: large sample size, many datasets & variables
- **Variety**: all possible data sources
- **Velocity**: continuous data collection
Informed participants

Demographics, behaviors
Lab-data, Bio-specimen
Clinical information, images
Healthcare service, Death

Questionnaire
Sensors
Research institutes
Research institutes
Medical institutes
NHI database
Death registry
Regional, Environmental Data

Cohort Database

Health-related Behaviors
Healthcare Service Utilization
Clinical, Laboratory, Omics

Primary researchers
Public health organizations
Secondary researchers

Limited linkage
Information Flow in Precision Medicine Cohorts

- Regional Data: Climate, Environment, Social, Economic...
- Informed individual participants:
  - Behaviors, life-log...
  - Genetic, Biosample...
  - Clinical, Image...
  - Medical use, death...
- Pre-existing cohort participants with additional informed consents
- Uploading with agreement:
  - Web, Mobile, Sensor, IoT...
  - Research center, Screening center, Biobank...
  - Hospitals, KCDC, KNHI, KHIDI...
  - NHIS, HIRA, NSO...
- Feedback of results with interpretation
- Interested researchers
  - Industry
  - Public
  - Secondary researchers

Cohort platform (Cloud)

- Life-log data
- Ecological data
- Medical utilization data
- Biological (omics) Data
- Public health data

Results
CONTENTS

- Precision Medicine (PM), PMI & PMI cohort
- Conventional cohort & PMI cohort
- Healthcare Big Data for Precision Medicine in Korea
Healthcare Big Data in Korea

- **Structured EHR Data**
- **Unstructured Clinical Notes**
- **Medical Imaging Data**
- **Genetic Data**: KoGES (>200K)...
- **Epidemiology & Behavioral**: Cohort studies, KNHANES...
- **Mass Screening**: National health screening, Private screening
- **National Health Insurance**: qualification, claims, review data
- **Registries**: Cancer, Severe diseases, Rare diseases
- **Vital statistics**: Causes of Death Database
EMR in a Korean Hospital – Yonsei University Health System

PACS 827 TB + OCS/EMR 17 TB for >6 million patients (As of 2015)
Medical Imaging Data in Korea

Number of CT, MRI, PET scanners

Data Source: Health Insurance Review & Assessment Service
Medical Imaging Data in Korea

Number of CT, MRI, PET examinations

Data Source: Health Insurance Review & Assessment Service
Hospital Data in Korea

- High rate of computerization
- Abundant imaging and laboratory data
- Earlier stage of EMR in most hospitals
- Insufficient & non-standardized clinical notes
Epidemiologic data

- The Korean Genome and Epidemiology Study (KoGES) Consortium

**KoGES**

- **Population-based studies**
  - KoGES_Ansan and Ansung study
  - KoGES_HEXA study
    - 17,337 (2004–2013)
  - KoGES_CAVAS

- **Gene-environment model studies**
  - KoGES_twin and family study
  - KoGES_immigrant study
    - 6,709 (Immigrants 2,875, Korean spouses 1,911, Children 1,923) (2006–2011)
  - KoGES_emigrant study
    - 3,563 (1,063 in Japan, 2,500 in China) (2005–2011)

- **Baseline**
  - 1st: 8,603 (2003–2004)

- **Follow-up**
  - 5th: 6,238 (2011–2012)

**N=225,396**
Epidemiological Data in Korea

- Increasing in quantity and quality
- Storage of biospecimen

- Relatively short history
- Decreasing participation rate
- Personal information and privacy issues
- Difficult to plan a long-term studies
Health Screening Programs in KOREA

- **Ministry of Health & Welfare (2015)**

<table>
<thead>
<tr>
<th>Programs by Ministry of Health &amp; Welfare</th>
<th>Subjects</th>
<th>Eligible</th>
<th>Participants</th>
<th>Participation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Health Screening</td>
<td>Age 40+, Employees of all ages</td>
<td>17,356,727</td>
<td>13,213,329</td>
<td>76.1%</td>
</tr>
<tr>
<td>Life Turning Point Health Examinations</td>
<td>Age 40, Age 66</td>
<td>1,052,609</td>
<td>812,082</td>
<td>77.1%</td>
</tr>
<tr>
<td>Health Screenings for Infants</td>
<td>4 to 71 months</td>
<td>3,134,526</td>
<td>2,178,441</td>
<td>69.5%</td>
</tr>
<tr>
<td>Cancer Screening</td>
<td>Stomach, Liver, Colon, Breast, Uterine cervix</td>
<td>20,179,798</td>
<td>9,749,039</td>
<td>48.3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>41,723,660</td>
<td>25,952,891</td>
<td>62.2%</td>
</tr>
</tbody>
</table>

- **Ministry of Education (2015):** 84,815 students
- **Ministry of Labor (2015):** 1,722,384 high-risk employees
- **Ministry of Defense:** unknown number of soldiers

- **Private Health Screening Programs:** >1 million per year
National Registries

National Cancer Registry
  • >200,000 cases per year
  • Accumulated 3.6 million cases (1988-2015)

Korea Acute Myocardial Infarction Registry (KAMIR-NIH)
  • N = 39,978

Korean Acute Heart Failure registry (KorAHF)
  • N = 5,625
Public Domain Data in Korea

- Increasing in quantity and quality
- High coverage rate
- Personal identification & Linkage between database

- Secondary (not research-purpose) data
- Standardization & validity issues
- Personal information and privacy issues
The integration between different domain data might be more important than the technological improvement in each domain.
THANK YOU FOR ATTENTION