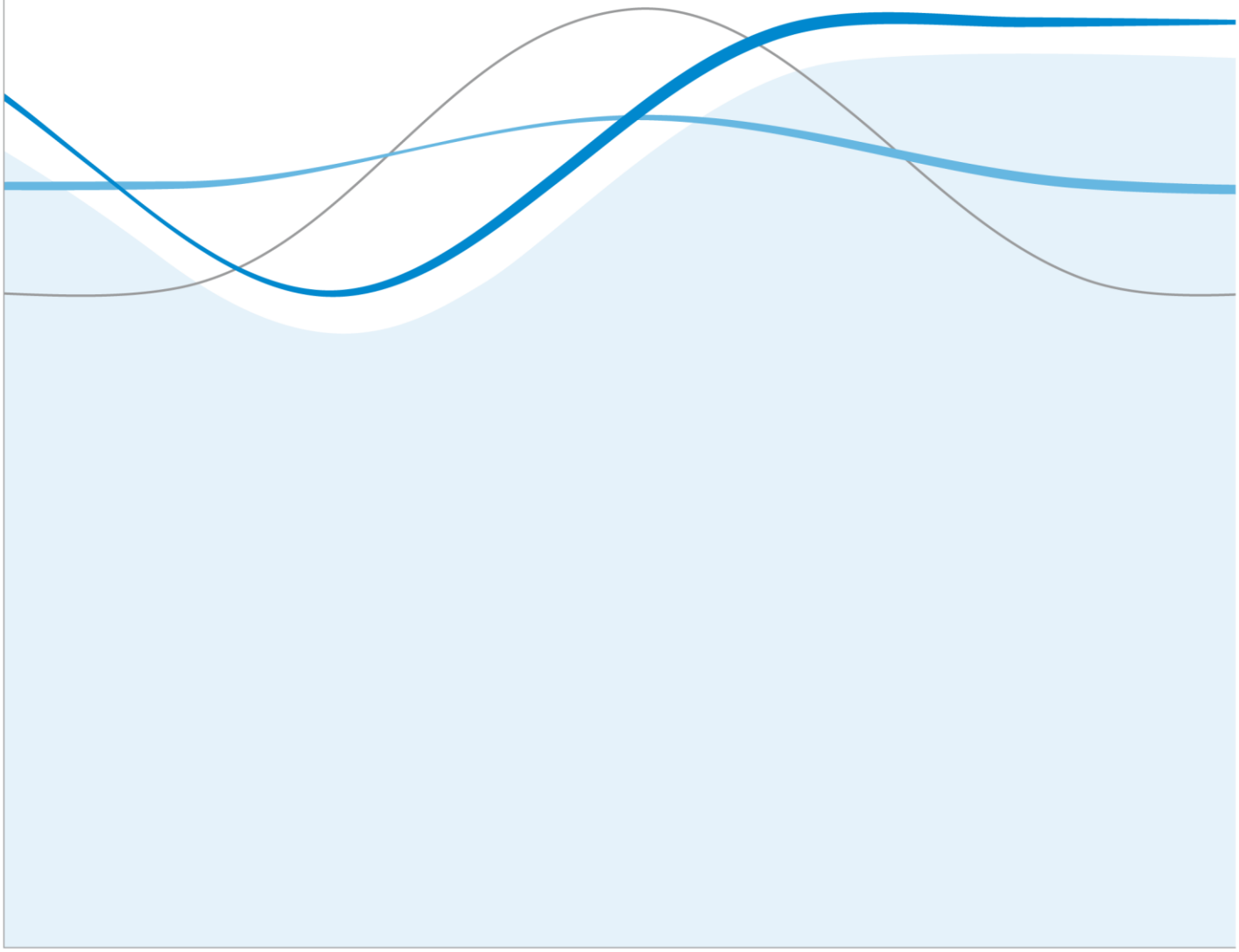
Office of Health Economics Consulting

Consulting Report

Data Governance Arrangements for Real-World Evidence: South Korea

Eui-Kyung Lee, Jae-A Park, Amanda Cole and Jorge Mestre-Ferrandiz

September 2017



Data Governance Arrangements for Real-World Evidence: South Korea

Eui-Kyung Lee[a], Jae-A Park[a], Amanda Cole[b] and Jorge Mestre-Ferrandiz[c]

[a]SungKyunKwan Universtiy (SKKU) [b]Office of Health Economics [c]Visiting Fellow, Office of Health Economics

September 2017

Submitted by:
OHE Consulting Ltd.
(a registered company number 09853113)
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom

For further information please contact:
Paula Lorgelly
Deputy Director and Director of Consulting
Tel: +44 (0)207 747 1412
Or: +44 (0)7789 435 855
plorgelly@ohe.org

About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE Consulting as OHE Consulting Reports are subject to internal quality assurance and undergo a rapid review by a member of the OHE Editorial Committee.

Acknowledgements

We would like to thank Lilly for commissioning and funding this project and for their permission to publish the report. We alone are responsible for any errors and omissions.

Table of Contents

1. INTRODUCTION

Following the publication of the OHE Consulting Report "Data Governance Arrangements for Real-World Evidence" (Cole et al., 2015), Lilly commissioned OHE Consulting to produce a report on "Data governance arrangements in South Korea". We have used the same method and structure as the original report.

Two of the authors (Eui-Kyung Lee and Jae-A Park) collected information via desk research through the same pre-specified pro-forma (which is included for completeness in Appendix 1). Amanda Cole and Jorge Mestre-Ferrandiz, in collaboration with Eui-Kyung Lee and Jae-A Park, summarised the information on South Korea in a tabular form (Table 6), using the same structure as Table 6 of the original report ["Data governance country comparison"]. The Table is structured along four main headings: "Data Protection – Health"; "Data Linkage"; "Access" and "Governance ideals and changes in the environment".

We also outline a "heat map" for South Korea (Figure 8), comparing South Korea's governance arrangements with the "ideal governance framework" suggested by Cole et al. (2015).

2. ARRANGEMENTS FOR DATA GOVERNANCE IN SOUTH KOREA

2.1. Brief overview of the health system & collection/management of patient data

2.1.1. Health system in South Korea

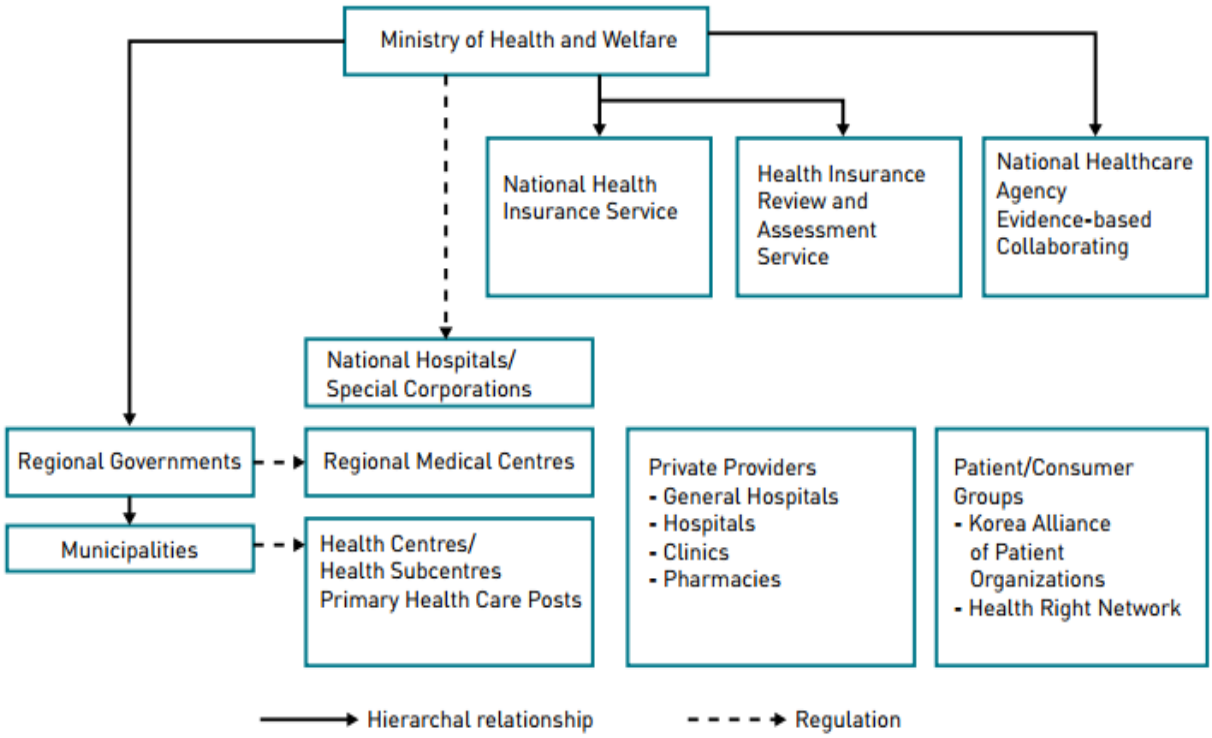
The South Korean health system achieved universal coverage in 1989. In 2000, the current health insurance system, which consists of a single payer with a uniform contribution schedule and benefits package, was established. Health care is provided for the entire population from either national health insurance, which covers 97% of the population, or the government-subsidised Medical Aid Program, which covers 3% of the population and provides healthcare services for low-income citizens. As one of the government organisations, the Ministry of Health and Welfare undertakes planning, establishing, and implementing national policies. Regional governments collaborate with the Ministry of Health and Welfare and manage regional health centres and facilities.

The National Health Insurance Service (NHIS) and the Health Insurance Review and Assessment Service (HIRA) are the two quasi-public organisations which manage the National Health Insurance (NHI) with the delegation from the Ministry of Health and Welfare. NHIS manages beneficiaries, collection of contributions, and payment to healthcare providers. It negotiates, with the representatives of different types of provider, the fee levels for the following year. HIRA reviews medical claims filed by healthcare providers for reimbursement and sends the results to NHIS, which then reimburses healthcare providers. It also establishes guidelines for the quality assessment of healthcare services. The National Evidence-based Health-care Collaborating Agency (NECA), which was established in 2009, is another quasi-public agency that executes health technology assessment. The results of its assessment of clinical effectiveness and cost-effectiveness of healthcare services, technologies, and products provide evidence

for consumers, healthcare providers, and health policy decision-makers (Kwon, Lee and Kim, 2015).

While the national health insurance is financed by the public sector, most of the healthcare providers belong to the private sector. They are designated as healthcare providers for the NHI beneficiaries and are not allowed to withdraw from designation. In other words, no matter which (public/private) sector a healthcare provider is in, it must provide healthcare to NHI beneficiaries. Their quality of care or services provided are assessed by HIRA. Figure 1 outlines the organisation of the health system in South Korea.

Figure 1. Organisation of the health system in South Korea



Source: Asia Pacific Observatory on Health Systems and Policies (Kwon et al., 2015)

2.1.2. Patient data collection

The single-payer health insurance system has allowed the collection of data from the entire population. Since all healthcare providers file reimbursement claims to HIRA and the results are then sent to NHIS, both HIRA and NHIS have claims databases. The purpose of maintaining these databases is to provide information on patients' utilisation of health resources for policy makers and public health researchers (Lee et al., 2016). These databases are very significant sources of real world evidence (RWE) as they contain a wide range of specific information on patients, from basic information to disease and healthcare resource utilisation.

2.1.3. NHIS data

NHIS has a claims database in which participants were followed for 14 years, from 2002 to 2015. From this database, the National Health Insurance Sharing Service (NHISS)

provides two types of data: National Sample Cohort (NSC), which is a population-based cohort, and customised databases that can be provided upon users request and assessment of its purpose. The NHIS database has a number of components/datasets: a sample cohort, health examinations, seniors, working women, and infant examinations. Each database contains the eligibility of the insured, medical treatments, medical examinations, and medical care institution. The structure of the database is shown in Table 1. In South Korea, everyone is issued with a unique resident registration number (identification number - like a social security number). To ensure anonymity of a patient, the patient's resident registration number is de-identified. NHIS thereby gives each patient an alternative serial-number ID (see Table 1), which is one of the de-identification methods used.

Table 1. Structure of the NHIS database

Database		Information
Eligibility of the insured		Basic information: alternative serial-number ID, sex, age group, location, insurance type, income group, disability status, etc.
Birth and death record		Alternative serial-number ID, sex, year of birth, year and month of death, first cause of death, second cause of death, etc.
Medical treatment	Specifications (20t*)	Alternative serial-number ID, key sequence number of claim, alternative medical care institution ID, date of first visit/treatment, medical department visited, primary diagnosis, secondary diagnosis, number of days treated or prescribed, number of hospitalisation days, costs incurred, etc.
	Treatment details (30t)	Key sequence number of claim, date of first visit/treatment, medical practice code, unit cost, total cost, daily dose, single dose, total prescription days, etc.
	Disease details (40t)	Key sequence number of claim, date of first visit/treatment, medical practice code, disease code, etc.
	Prescription (60t)	Key sequence number of claim, date of first visit/treatment, treatment type code, prescription data (generic name, single dose, daily dose, total days prescribed, etc.), unit cost, total cost, etc.
Medical examination		Health examination data: year of examination, alternative serial-number ID, examination institution code, height, weight, waist, high blood pressure, low blood pressure, total cholesterol, HDL cholesterol, LDL cholesterol, creatinine, AST level, ALT level, disease history, family disease history, cancer status, diabetes status, smoking status, smoking status, exercise status, etc.
Medical care institution		Medical institution data: institution's alternative serial-number ID, institution type code, location, number of beds, number of doctors, CT status, MRI status, PET status, etc.

*Note: NHIS named each type of dataset as 20t, 30t, etc., which is similar to how HIRA named its datasets as Table 20, 30, etc. (see Table 2). For example, when someone refers to dataset 20t, it means the 'medical treatment – specifications' dataset was used.

From the NHIS database, various useful information can be elicited. For instance, the user can analyse mortality of the 'insured' database by evaluating date of death, or loss of eligibility, as the participant becomes ineligible only upon death given that health insurance is universal in South Korea. The user can also comprehend insured participants' economic status from insurance premium data, and use clinical data from medical examinations. These can be used for evidence-based decision making and public policy.

2.1.4. HIRA data

Similar to NHIS, HIRA provides customised datasets and cross-sectional sample data of insurance claims filed by healthcare service providers from its database. Customised datasets can be provided upon users request and assessment of its purpose. These data are open to the public, including members of the private sector, with a charge. A study proposal is required. This means that, unlike NHIS data which is restricted to those with academic or public policy affiliations, HIRA data can be provided to the members of private sector, such as pharmaceutical or consulting companies. The structure of HIRA's data is shown in Table 2.

Table 2. Structure of HIRA's database

Database	Information
Specifications (Table 20)	Age, gender, type of national health security enrolled, sampling weight of the patient, medical department visited, primary diagnosis, secondary diagnosis, alternative serial-number ID, key sequence number of claim, etc.
Treatment details (Table 30)	Medical practice code, incurred costs, key sequence number of claim, inpatient prescription data (generic name, daily dose, unit cost, total days prescribed), etc.
Disease details (Table 40)	All diagnoses and key sequence number of claim
Prescription (Table 53)	Medical practice code, incurred costs, key sequence number of claim, inpatient prescription data (generic name, daily dose, unit cost, total days prescribed), etc.
Medical care institution (Table yno*)	Institution's alternative serial-number ID, location, number of beds for health care institutions, etc.

*Note: The name of the table for 'Medical care institution' is "Table yno.", rather than assigning it a number (as with other tables).

The HIRA national sample data includes four samples: the National Patients Sample (NPS), the National Inpatient Sample (NIS), the Adult Patient Sample (APS), and the Paediatric Patient Sample (PPS). These are sample datasets that are readily available upon completion of an application and approval of provision. Unlike customised datasets (which must be processed according to the applicant's requests) these sample datasets are already processed and can be readily provided, once HIRA approves data provision (following submission of study protocol, institutional review board (IRB) approval notice, a charge, etc.; this process is explained in more detail below). A description of each of the four sample datasets is shown in Table 3.

Table 3. Type and description of HIRA national sample data

Type	Characteristics	Available data
NPS	- All patients - 3% of all patients (approx. 1.4 million)	2010-2015
NIS	- Mostly inpatients - 13% of inpatients (approx. 0.7 million) - 1% of outpatients (approx. 0.4 million)	2009-2015
APS	- Elderly patients - 20% of elderly patients: ≥ 65 years old (approx. 1.0 million)	2010-2015
PPS	- Paediatric patients - 10% of paediatric patients: < 20 years old (approx. 1.1 million)	2010-2015

2.1.5.Access to claims data

The NHIS and HIRA databases are mainly open to researchers with academic or public policy purposes, with charge. However, there are minor differences between the access policies for NHIS and HIRA databases, with the former being more restrictive.

For the NHIS database, requests for claims data to undertake studies with a commercial interest are not allowed. Under NHIS data provision operation regulation article 12, eligible users are as follows: 1) central administrative agency or local government agency, public institution or public research institute; 2) persons carrying out research supervised by such institutes from 1; 3) persons affiliated to an institute from 1 that are carrying out research for publication purposes; 4) persons carrying out research for dissertations and; 5) persons carrying out research for other public purposes.

Data are provided based on the eligibility of the principal investigator of the corresponding research. In the case of a principal investigator receiving financial support from a private sponsor (e.g. a pharmaceutical company), the eligibility standard 3 can be applied only for publication purposes and if the research report is not provided to the corresponding private sponsor.

In case a pharmaceutical or consulting company wish to request data access, the eligibility standard 5 applies. In this case, data are not provided directly to the company, and the researcher must visit NHIS to access the sample database through a computer in the analysis center. Using the database for commercial purposes is strictly prohibited. Furthermore, customized databases are not subject to provision for private companies. Research contents are assessed for public interest through preliminary review of a research proposal and the results are also assessed for their scope, which must be within the public interest, before being provided to the company.

The HIRA database is mainly open to researchers with academic or public policy purposes as well. If the purpose of the research is academic (mainly for publication),

researchers from academia can use the HIRA database even when they are receiving financial support from a private sponsor.

On the other hand, pharmaceutical companies can get limited access to the database. Data for their own drug products are provided, and other companies' drug products can be provided only with their consent. Furthermore, data are provided with up to thirty variables. For data other than their own products, information is given as INN (International Nonproprietary Name) and therapeutic classes (ATC codes).

For both HIRA and NHIS databases, the proposal for database request is examined by a Review Committee.

2.1.6. Electronic Medical Record data from healthcare institutions

In addition to the NHIS and HIRA databases, many healthcare providing institutions (such as hospitals) collect patient information via an Electronic Medical Record (EMR). Whilst this presents a strong potential source of real world data (RWD), the EMR system is not mandatory for healthcare providers, and there are no standardised rules that cover all providers. There is insufficient information about requesting EMR data use, and there is a need to contact each healthcare provider for information on this. There is no information on the coverage of the EMR system among the South Korean population.

Compared to insurance claims data, EMR offers richer clinical and cost information of patients, which allows better control of confounding factors during analyses. Also, it captures healthcare services that are not covered with health insurance. However, it does not capture healthcare utilisation outside the institution. Recording patient data needs a clinician's involvement, which might affect the quality of data. Thus, quality of data will vary among healthcare institutions, affecting the feasibility and validity of analyses.

2.2. Core legislation and governance arrangements for the collection and/or use of patient data

2.2.1. Routinely collected patient data

Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

Increasing privacy risks due to technology development requires strong regulations on the protection of personal information. In order to allow safe use of big data within the current legal framework, the Office for Government Policy Coordination, the Ministry of the Interior, the Korea Communications Commission, the Financial Services Commission, the Ministry of Science, ICT and Future Planning, the Ministry of Health and Welfare, and other related agencies collaborate and provide guidelines on the protection and de-identification of personal information, as well as what it can be used for.

Since both NHIS cohort data and HIRA data are open to the public, they are considered as open data and must be used under the legal framework for using personal confidential data, including: Act on Promotion of the Provision and Use of Public Data 2013 (PUPD) and Personal Information Protection Act 2011 (PIPA). These two Acts are (briefly) described below.

2.2.1.1. Act on Promotion of the Provision and Use of Public Data 2013 (PUPD)

The purpose of PUPD is to prescribe matters for promoting the provision and use of data held and managed by public institutions in order to guarantee citizens' right to access public data, to contribute to improving their quality of life, and to develop the national economy through the utilization of such public data in the private sector. Under the PUPD, public institutions that hold, manage, and provide data to the public must adhere to the provision and use of public data principles, which cover:

- enabling anyone to readily use public data and taking measures necessary to promote universal access to the use thereof,
- guaranteeing citizens equality in their access to, and use of, public data,
- not impeding the use of public data disclosed to the general public through information and communications networks,
- not prohibiting or restricting the use of public data for gain,
- compelling every use to observe obligations prescribed under statutes and the terms and conditions of use to prevent any violation of public interests.

Whilst PUPD emphasises enabling the broad utilisation of public data, NIHS and HIRA must also consider and implement the regulatory framework around the protection of personal information, described below.

2.2.1.2. Personal Information Protection Act 2011 (PIPA)

The purpose of this Act is to prescribe matters concerning the management of personal information in order to protect the rights and interests of all citizens and further realize the dignity and value of each individual by protecting personal privacy from collection, leakage, misuse and abuse of individual information. Under PIPA, a personal information manager that manages personal information must adhere to the personal information protection principles, which cover:

- making clear the purpose of managing personal information, collecting personal information lawfully and legitimately, and limiting the collection to the minimum extent necessary to achieve such purpose,
- managing personal information within the appropriate extent necessary for achieving the purpose of managing the personal information, and not using it for the purposes other than intended ones,
- guaranteeing that personal information is kept accurate, complete, and up-to-date to the extent necessary for the purpose of managing the personal information,
- managing personal information safely, in consideration of the risk that the rights of a subject of information may be violated and the level of accompanying risks depending, among other things, on the management methods and kinds of personal information,
- disclosing to the general public matters concerning the management of personal information, including, but not limited to, personal information management policies, and guaranteeing the rights of a subject of information such as the right to request an inspection,
- managing personal information in such a manner that the privacy infringement of a subject of information is minimised,

- ensuring that personal information is managed anonymously whenever such management is possible,
- endeavouring to gain the trust of a subject of information by fulfilling his/her responsibilities and obligations conferred or imposed by or under this Act, relevant Acts and subordinate statutes.

Since the NHIS NSC database and the HIRA national sample data both contain personal information, they must undergo de-identification. The process of de-identification follows a four-stage procedure:

1. Preliminary review
2. De-identification
3. Propriety assessment
4. Post management

2.2.1.3. Preliminary Review

In order to handle patient data for big data analyses, it must be preliminarily reviewed for personal information. If the data is deemed as not personal information, it can be used for big data analyses without additional actions. If the data is deemed as personal information, it must be processed through de-identification.

2.2.1.4. De-identification

For patient data considered as personal information, the identifier must be removed. The identifier is a unique name or value given to a person or related object. If the identifier is needed for data use or analysis, it must be de-identified first (e.g. resident registration number must be anonymised, for example by creating an alternative serial-number ID). Examples of identifiers are: resident registration number, passport number, driver's license number, name, specific address, telephone number, bank account number, and e-mail address. Attribute values must be removed as well, unless they are related to the purpose of data use or analysis. Attribute values are the information that allows identification of a person when combined with other information. If attribute values are needed for data use or analysis, they must be de-identified through techniques such as pseudonymisation, aggregation, data masking, data suppression, and data reduction. Each technique uses various specific skills and should be adopted considering the appropriateness and the purpose of data use.

2.2.1.5. Propriety Assessment

When personal information is not completely de-identified, the concern about possible identification through inference method or data combination arises. Thus, de-identified data must undergo propriety assessment and the procedure is as follows.

1. Preparing base data – For propriety assessment, the personal information processor needs to fill out base data, including data details, current state of de-identification and data user's level of management.
2. Forming the evaluation group – Three or more personal information protection managers form the evaluation group.

3. Assessment – The evaluation group assesses the propriety of de-identification using the base data and k-anonymity model[1].
4. Additional de-identification – The personal information processor must perform additional de-identification that reflects the evaluation group's opinion if the assessment result shows impropriety.
5. Data use – If the assessment results show propriety, use or provision of data for analyses is allowed.

When de-identification of personal information is completed, HIRA national sample data and NHIS NSC databases can then be used.

2.2.1.6. Internal Instructions and Guidelines

NHIS and HIRA both have internal guidelines on provision and use of public data, which are, respectively, the Operating Regulations on National Health Information Data Provision and the Operation Guideline on Public Data Provision and Use. A brief description of both is shown in Table 5, and details can be found on their websites.

2.2.2. Collecting de novo patient data

Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

In Cole et al. (2015) a distinction is made between the governance arrangements around the use of routinely-collected patient data and the collection of de-novo patient data (i.e. the process and rules around collecting new information from patients). The distinction is mainly due to the fact that the purpose and scope of new data-collection activities are often subject to review, and patient consent can generally be sought more readily. In Korea, there is no additional information on the collection of de novo patient data. Since both NHIS and HIRA data are reimbursement claims from healthcare providers, claims data of all patients covered by NHI and Medical Aid Program are included.

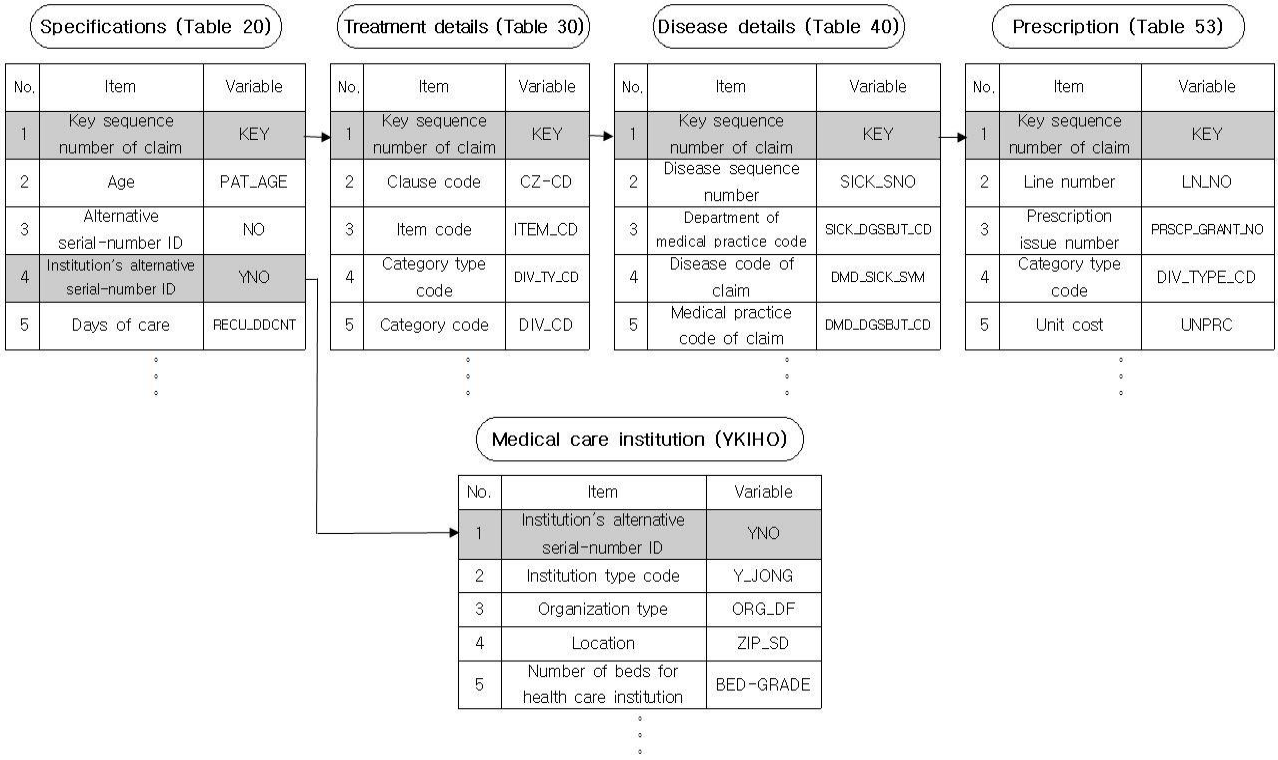
2.3. Data linking

To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

Patient data can be linked across datasets only within the same database. There is no ability to link HIRA data to NHIS data or electronic records. For instance, datasets within HIRA's national sample data can be combined by linking a common variable, such as the 'key sequence number of claim'. Referring to Table 2 above, which sows the structure of HIRA's national sample data, '*Table 20*' (which refers to 'Specifications') can be linked with *Tables 30, 40, and 53* via the key sequence number of claim. A descriptive diagram of such links is shown in Figure 2.

[1] The k-anonymity model is one of the privacy protection models used for propriety assessment.

Figure 2. Linkage within HIRA's National Sample Data



Similarly, datasets within each NHIS NSC database can be combined. For instance, the datasets 20t, 30t, 40t, and 60t within the medical treatment database (described in Table 1 above) can be linked via the key sequence number of claim. A descriptive diagram is shown below in Figure 3.

What distinguishes NHIS NSC data from HIRA data is that NHIS also provides eligibility and medical examination data of the insured patients, which allows further analysis through data combination across different datasets. The combined medical treatment dataset of 20t, 30t, 40t, and 60t can be linked with both the 'eligibility of the insured' and 'medical examination' database via the patients' alternative serial-number ID. In addition, the 'medical care institution' dataset can be combined via the institution's alternative serial-number ID for medical care institutions. A descriptive diagram is shown below in Figure 4.

Figure 3. Linkage within the NHIS NSC database

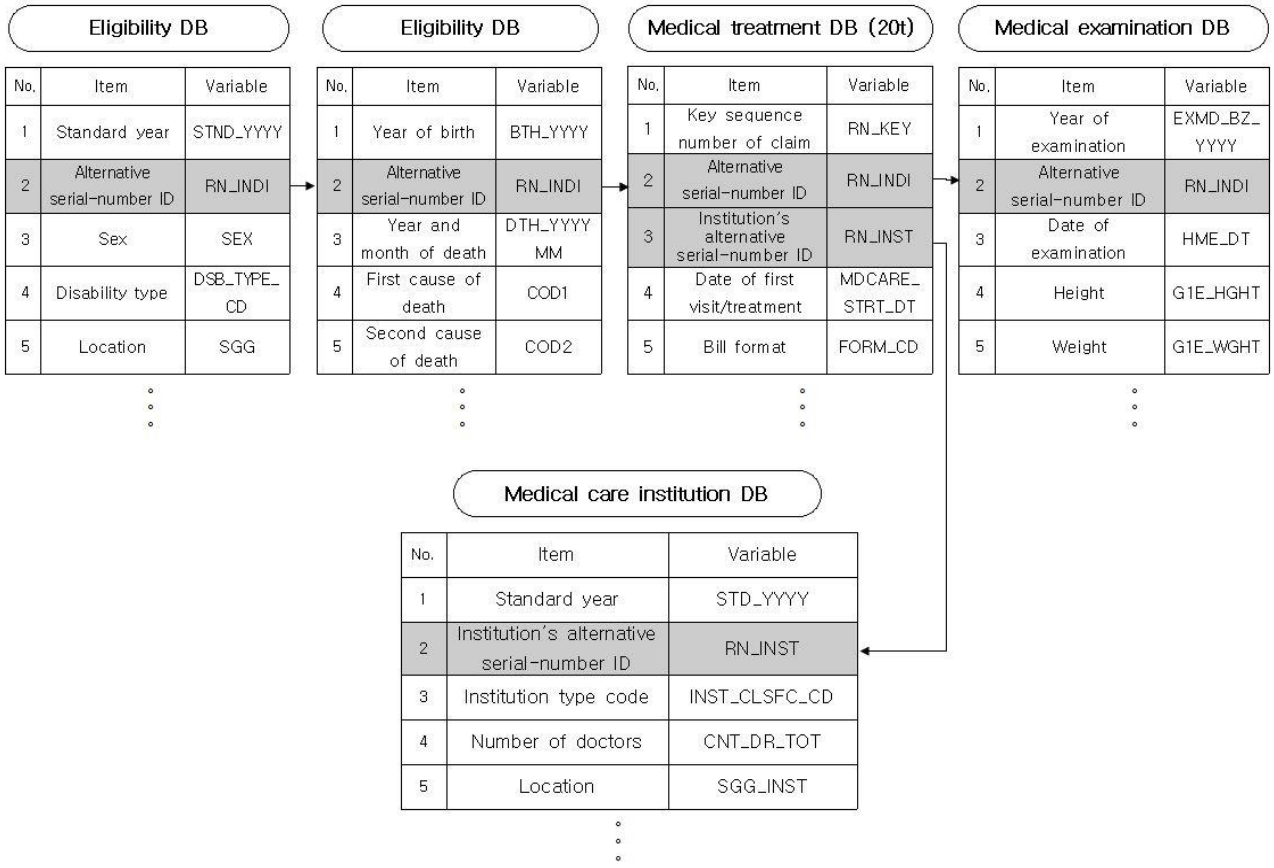
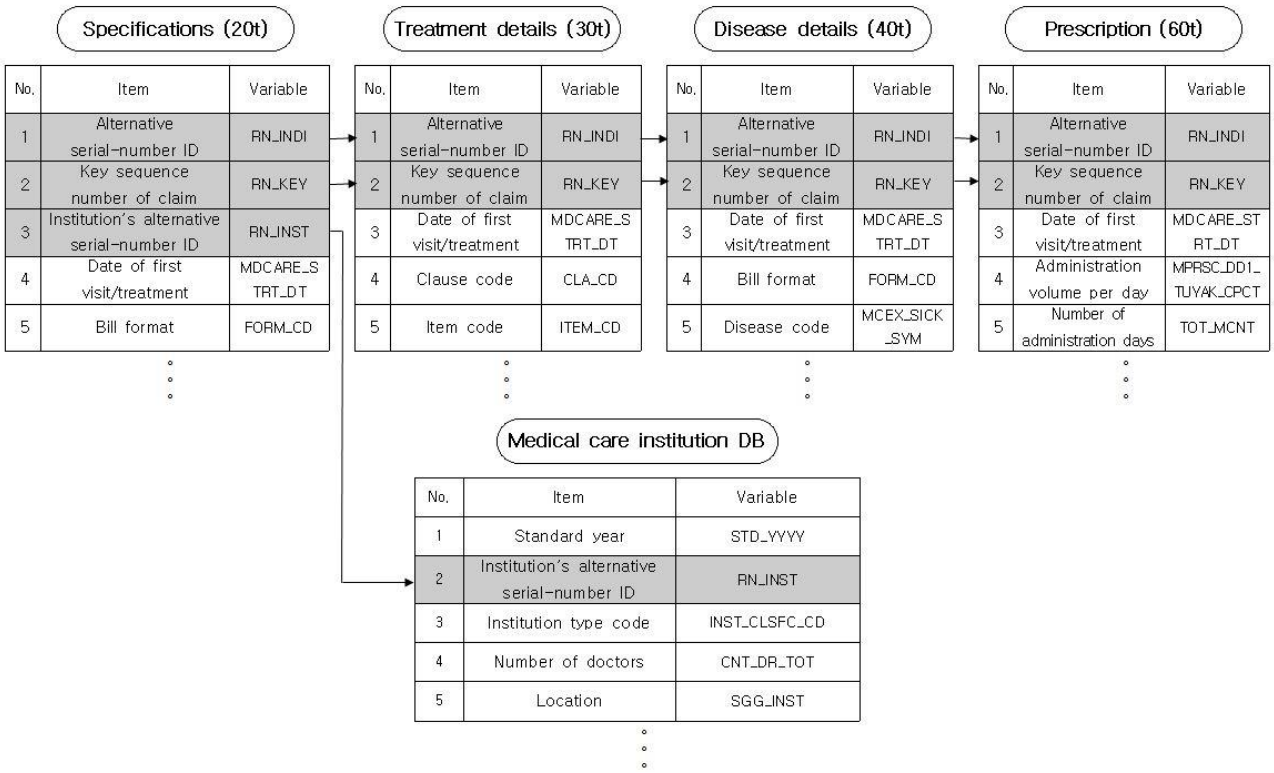


Figure 4. NHIS data combination across different databases



2.4. Data access

To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

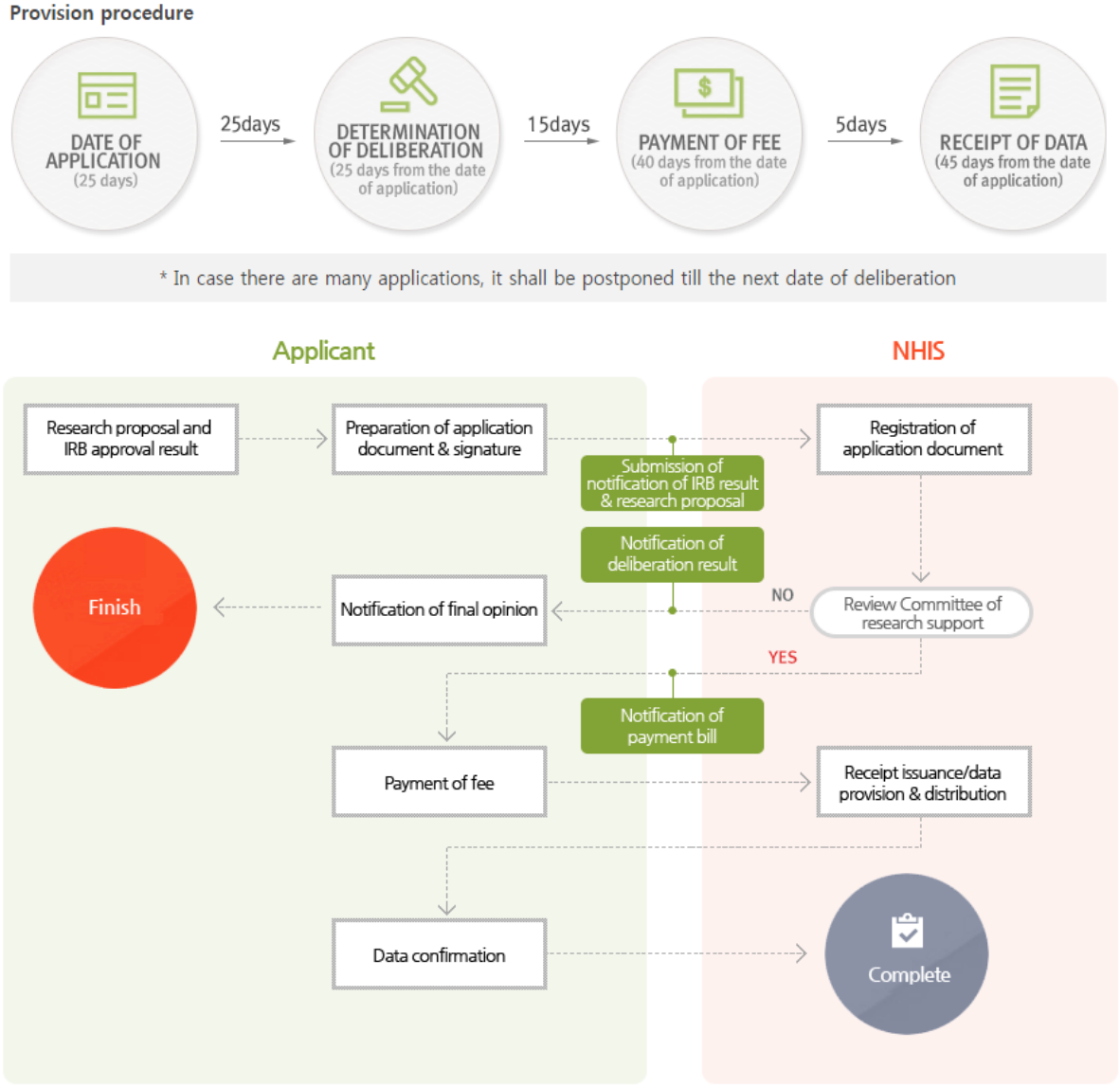
As mentioned above, both NHIS and HIRA provide customised and sample databases for those with policy and academic research purposes. In addition, HIRA provides claims data to members of the private sector as well.

2.4.1. NHIS data access

NHIS offers both sample and customised data of health insurance claims filed by medical care institutions. For a sample dataset, it takes a total of 45 days to receive the data through three stages of provision, including determination of deliberation (25 days), payment of fee (15 days), and receipt of data (5 days). Applicants must submit a research outline, a research plan, and an approval notice from an institutional review board (IRB), registering the necessary application documents online. The duration of the process depends on each institution's IRB. The Review Committee assesses the propriety of provision and notifies the applicant of the result. If the application is accepted, the invoice will be sent. After payment is completed, the applicant will receive datasets from NHIS via mail. Figure 5 below sets out the provision procedure and application guide in detail.

In addition to the sample database, NHIS also offers customised datasets. Potential users need to consult about having access to the customised data by visiting NHISS or over the telephone. After receiving the user's application, the Review Committee deliberates on the propriety of data provision. If the application is accepted, the data analysis department of NHIS extracts data and prepares the provision procedure. After security measures of data are handled and completed, NHIS provides data to the user and performs follow-up management. Figure 6 outlines the operation guide to apply for customised datasets.

Figure 5. Provision procedure and application guide to access NHIS sample datasets

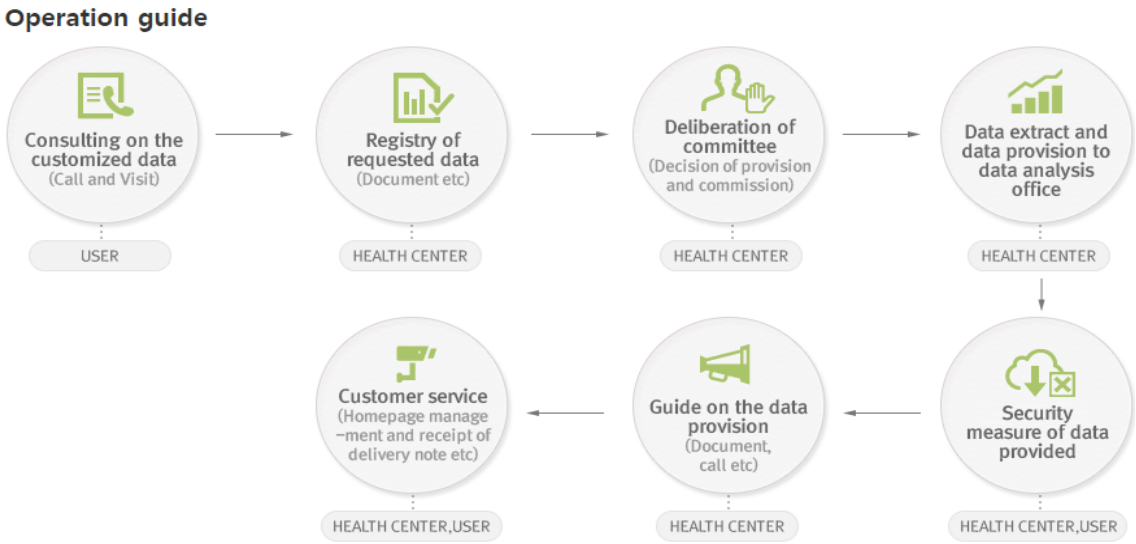


Source: National Health Insurance Sharing Service

NHIS has two big data analyses centres for analysis and provision of public data, located in its headquarter Seoul office. Data users can visit these centres and operate their analyses on site. While sample data can be provided in compact discs via mail, customised data must be accessed on site.

Applicants can find application information and materials on NHISS website (https://nhiss.nhis.or.kr). Table 4 shows the subject criteria, time period, and contents for five databases: the sample cohort (NHIS-NSC), the medical check-up cohort (NHIS-HealS), the elderly cohort (NHIS-Senior), the working women cohort, and infant medical check-up cohort. Further details can be found on the NHISS website (https://nhiss.nhis.or.kr/bd/ab/bdaba002cv.do).

Figure 6. Guide to apply for customised NHIS datasets



Source**:** National Health Insurance Sharing Service

Table 4. Structure of NHIS Sample Research databases

Type	NHIS-NSC	NHIS-HealS	NHIS-Senior	NHIS-Working Women	NHIS-InfantHealS
Subject Criteria	The eligible subjects as of 2006 (approximately 1 million people)	Regular health examinees between the age of 40 and 79 in 2002-2003, with eligibility as of 2002 (approximately 510 thousand people)	Subjects over 60 years of age, with eligibility as of 2002 (approximately 550 thousand people)	The eligible subjects as of 2007 who are working women in the age of 15-64 (approximately 180 thousand people)	Out of total health examinees who received at least one of 1st or 2nd medical check-up, 5% sample is extracted for each birth year between 2008 and 2012
Time Period	2002-2015 (14 years)	2002-2015 (14 years)	2002-2015 (14 years)	2007-2015 (9 years)	2008-2015 (8 years)
Contents	Socioeconomic data including disability and death, medical care history (medical treatment and health examination), and medical care institution	Socioeconomic data including disability and death, medical care history (medical treatment and health examination), and medical care institution	Socioeconomic data including disability and death, medical care history (medical treatment and health examination), medical care institution, and status of long term care services	Socioeconomic data including disability and death, medical care history (medical treatment and health examination), and medical care institution	Socioeconomic data including disability and death, medical care history (medical treatment and health examination), and medical care institution

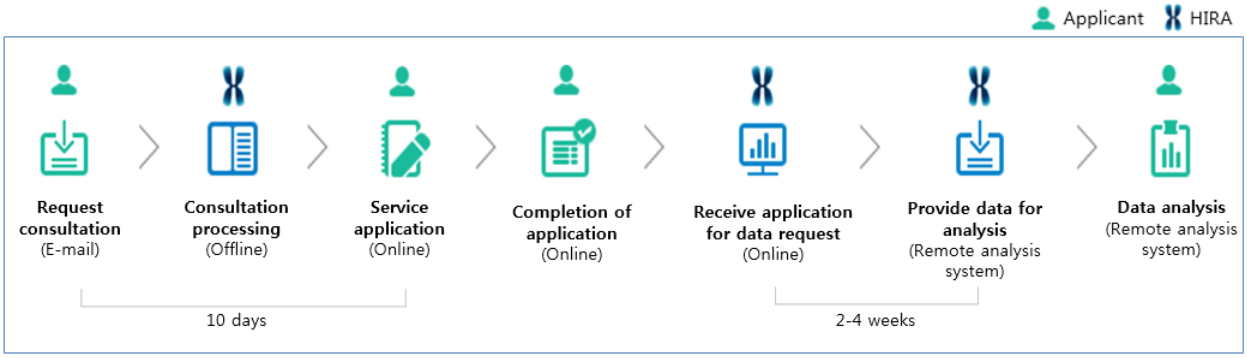
2.4.2. HIRA data access

HIRA provides both sample and customised datasets when the applicant submits necessary information and forms. Based on health insurance claims data, a sample dataset provides healthcare utilisation information on patients receiving treatment for each year. Below are conditions on data provision.

- The patient dataset will be sent after payment has been confirmed.
- The patient dataset cannot be used as evidence of policy-making because there are sample errors of sample data.
- Applicants must submit information on affiliated institute, level of education of data user, and research title.
- A large statistical tool is needed to handle these data sets.

The application process for both sample and customised public data of HIRA is very similar to that of NHIS. For sample data, the applicant must submit an application online and have a consultation with HIRA. After payment is made, HIRA provides sample data in compact discs via mail. For customised data, the applicant must request for consultation first and then submit an application online. Upon submission, the HIRA Review Committee assesses the application for propriety. If the application is accepted, HIRA provides customised data to the applicant. The application and provision process is shown in Figure 7, with timelines (where available).

Figure 7. HIRA Data application and provision process



HIRA established eight healthcare big data centres for analysis and provision of public data in various cities of South Korea. Data users can visit these centres and operate their analyses on site. Data users can also use their devices by connecting to the analysis server after authentication of the Review and Assessment Service Centre to operate big data analyses, via a remote access system. To export the results of the analysis, the user must submit an application for data export. Users must also submit analysis outputs to the director of HIRA within 30 days from the date the outputs occurred, including research reports and studies. The user must indicate their use of public data when the analysis is completed. There are fees for the provision of public data (see Table 5), but a user executing analyses for national and local governments is exempt from the fees and charges.

2.5. Data use

What, if any, are the rules governing the use of RWD?

2.5.1. Data recipients

NHIS and HIRA have been providing a variety of big data to researchers. NHISS provides the national health information data for state, local government, public institutions, institutions or person that conduct research in accordance with the agreement (MOU) signed with the Authority. HIRA data can be provided to people from private organisations as well.

In order to use public health information for research with academic or public policy purpose, the research group needs to designate an individual with domestic nationality residing in the Republic of Korea as a research officer.

2.5.2. The Review Committee: NHIS

The propriety of providing NHI data is assessed by the Review Committee. This Committee reviews the information such as details on the rationale for provision, period of data provision, details on objection and modifications, payment, exemption, and further details on research assistance. The Review Committee consists of seven to nine members, including a Chair. The Chair is the Director of the Big Data Steering Department, and can appoint up to two external experts. The Committee members include the Head of the Statistics Analysis Division in the Health Insurance Policy Research Institute, the Head of the Convergence Technology Division in the Big Data Steering Department, four employees from the Health Insurance Policy Research Institute and the Big Data Steering Department, and external experts if needed.

The Chair convenes the Review Committee twice a month. Meetings of the Committee require the attendance of the majority of the members, and pass resolutions with votes from the majority of those present.

2.5.3. The Review Committee: HIRA

HIRA also manages a Review Committee to deliberate on matters related to the provision of public data. The Review Committee consists of less than six members, and the Director of the Public Data Provision Department is appointed as the chairperson. The Review Committee manages the master plan and implementation plan of data provision, and establishes specifics on the assessment of the execution performance on the key measures about provision and use of the public data. Table 5 below shows the definition, data recipients, cost, and the processing of the sample and customised databases from NHIS and HIRA.

Table 5. Details on Data Use for NHIS and HIRA (Sample and customised data)

	Organisation	**NHIS**	**HIRA**
Sample Data	Type	- NHIS-NSC - NHIS-HealS - NHIS-Senior - NHIS-WorkingWomen - NHIS-InfantHealS	- National Patient Sample (NPS) - National Inpatient Sample (NIS) - Adult Patient Sample (APS) - Paediatric Patient Sample (PPS)
	Definition	Sample data for provision is standardised by extracting data from the database for the purpose of research	Based on the health insurance claims data, sample data of a patient, with a research purpose, is extracted for one year from the date of first treatment
	Data recipient	<**Operating Regulations of National Health Information Data Provision** (Article 4)> (1) National and local government (2) Public institution according to Article 2, Paragraph 6 in the <Technology Transfer and Commercialisation Promotion Act>. (http://elaw.klri.re.kr/kor_mobile/viewer.do?hseq=31892&type=part&key=28) (3) Institution or a person who conducts research for the agency that has a contract with an institution in accordance with No. 1 or No. 2 (4) Institution or a person who conducts research in accordance with the agreement (MOU) signed with public corporation (5) Other individuals who perform research in academia or public policy	<**Act on Promotion of the Provision and Use of Public Data** (Article 3, General Principles)> (1) Public data that are produced, possessed, and managed by HIRA shall be available to all citizens and be accessible and usable to citizens. (2) Public data that is provided to the public cannot prohibit or restrict the available public data, unless there are special provisions in other laws or public, regardless of the commercial and non-commercial purpose of use, with the exception of the case subparagraphs of Article 11 and Article 12
	Cost*	Provision Fee + Discharge cost (1) Provision fee according to number of days of use - < 7 days: ₩25,000/day - From 1 week to 1 month: [₩112,500/week + ₩25,000/remaining day]	Each patient dataset for ₩300,000 per provision

		- Greater than 1 month: : [₩350,000/month + ₩112,500/ remaining week + ₩25,000/remaining day ₩25,000/remaining day] (2) Discharge cost (USB): ₩10,000/GB	
	Operation guide	(1) Research proposal and IRB approval (2) Preparation of application document and signature (3) Registration of application document (4) Review Committee of research support (5) Notification of payment bill (6) Payment of fee (7) Receipt insurance/date provision and distribution (8) Data confirmation (9) Complete	Applicants can request sample through HIRA webpage. (http://opendata.hira.or.kr/op/opc/selectPatDataAplInfoView.do)
Customised Data	Definition	Customised data is provided as requested for researchers with academic or policy-related purposes	<**Act on Promotion of the Provision and Use of Public Data** (Article 2 Paragraph 1, Paragraph 2)> The term "public data" means any data or information, including databases and electronic files, processed in optical or electronic form, and created or acquired and managed by any public institution for the purposes set forth in statutes
	Data recipient	<**Operation regulation of National Health Information Data Provision** (Article 4 Paragraph 1)> (1) National and local governments (2) Public institution according to Article 2 in <Technology Transfer and Commercialisation Promotion Act> (3) Institution or a person who conducts research for the agency that has a contract with an institution in accordance with No. 1 or No. 2 (4) Institution or a person who conducts research in accordance with the agreement (MOU) signed with a public corporation (5) Other individuals who perform research in academia or	<**Act on Promotion of the Provision and Use of Public Data** (Article 3, General Principles)> (1) Every public institution shall endeavour to enable anyone to readily use public data and impede the use of public data disclosed to the general public through information and communications networks (2) No public institution shall prohibit or restrict the use of public data for gain, except as otherwise expressly provided for in any other Act or except in cases referred to in the subparagraphs of Article 11 paragraph 1 and Article 12 paragraph2

		public policy - If requested with regard to policy performance in a state agency - If requested from a general or university institute for the government - If an individual or organisation that requests to promote public welfare and other healthcare	
	Cost*	Provision and Data use fees (1) Provision fee according to number of days of use - < 7 days: ₩50,000/day - From 1 week to 1 month: [₩225,000/week + ₩50,000/remaining day] - Greater than 1 month: : [₩700,000/month + ₩225,000/ remaining week + ₩50,000/remaining day ₩25,000/remaining day] (2) Data use fee: ₩10,000/GB if greater than 200GB +	Number of days × fee per day (1) Fee per day: ₩50,000 (from the first day of analysis to the last day of analysis) (2) Exemption: State agencies and local governments, except affiliated organisations (3) Subject to 50% discount: MOU institutions
	Operation guide	(1) Consultation on the customised data (call or visit) (2) Application for customised data (3) Deliberation of Review Committee (4) Data extraction and provision by data analysis department of NHIS (5) Security measure of data provided (6) Guide on the data provision (documents, call, etc.) (7) Post management	(1) E-mail application consultation (research outline and proposal required) (2) Online application (3) Application process completion (4) Applications received (data request) (5) Data provision (remote analysis system) (6) Data analysis (remote analysis system)

Note: *: ₩1 (South Korean Won) = 0.000675027GBP; ₩1 (South Korean Won) = €0.000763079 (Source: www.xe.com; accessed 10 July 2017)

2.6. Governance ideals and changes to the environment

Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

No national documentation was identified that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

3. GOVERNANCE ARRANGEMENTS IN SOUTH KOREA AND COMPARISONS WITH OTHER COUNTRIES

The original OHE report (Cole et al., 2015) detailed the governance arrangements for RWD in Australia, France, Germany, Italy, Sweden, the Netherlands, the UK and the US. A table was provided summarising data protection, data linkage, access and governance ideals. In order to facilitate a direct comparison, Table 6 below summarises this headline information for South Korea.

The collection and use of routinely collected reimbursement data is well developed in South Korea, and the process for access to datasets is well defined, being more restrictive for NHIS data (restricted to researchers in public policy and academics) and more open for HIRA data to organisations or individuals outside of the public policy / academic sector (e.g. private organisations and pharmaceutical companies).

A common framework observed in other countries assessed for their real-world data governance arrangements (Cole et al., 2015) was one of 'consent or anonymise': i.e. that patient consent must be collected for patient data to be used for research, or if not then data must be completely anonymised (with most countries having provisions and processes in place to make data accessible where consent is not possible, in cases where that research would be of benefit to society). However, this does not appear to be the case in South Korea, and the issue of consent does not appear to be prominent in the governance arrangements in place. Rather, the framework in Korea encompasses the anonymisation (de-identification) of personal data. Whilst this satisfies data protection issues, it may limit the possibilities to link datasets and the resulting insight this could provide. There is no separate process regarding patient consent, since claims data are data that are already collected by healthcare providers for reimbursement purpose.

In Cole et al. (2015), an 'ideal governance framework' is suggested. In the heat map below, set out in Figure 8, we outline how the governance arrangements in South Korea compare with these ideals.

Table 6. Data governance arrangements, South Korea: Summary

Data Protection – Health [Patient consent & exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
The Act on Promotion of the Provision and Use of Public Data 2013 (PUPD) was introduced to promote and regulate the access to and use of public data, and covers: enabling anyone to readily access data, guaranteeing equality of access, not prohibiting the use of public data for gain, and ensuring conditions are met to prevent violation of public interests. The Personal Information Protection Act 2011 (PIPA) concerns the protection of citizens' privacy and personal information. A personal information manager must adhere to personal information protection principles including: - Having a clear purpose for collecting and managing personal data, and limiting collection to the minimum extent necessary to achieve such purposes - Guaranteeing information is accurate, complete and up to date - Managing the risks of personal information and minimising privacy infringement - The right for subjects to inspect data that is collected on them - Ensuring information is managed anonymously wherever possible Routinely collected patient data is mainly through health insurance claims databases through HIRA and NHIS, which contain patient	Personal information must be removed before use. Where an identifier is needed for data use or analysis, it must be de-identified prior to use. Identifiers can include: resident registration number, passport number, driver's license number and name. Generally, the alternative serial-number ID is used as a de-identified alternative to the resident registration number, and can be used for data linkage purposes, but only between datasets within the same database. De-identification is through pseudonymization, aggregation, data masking, data suppression or data reduction. Where personal information is not completely de-identified (could be inferred through inference or data combination), data must undergo propriety assessment (involves a formal	Both NHIS and HIRA provide customised and sample data for those with policy and academic research purposes, with a charge. The main sample databases are: the NHIS National Sample Cohort data (2002-2015) and HIRA national sample data (2010-2015). Customised databases can be provided upon a user's request and assessment of its purpose. Both organisations have review committees which consider applications. The applicant must assign a research officer to the project with domestic nationality and residing in the Republic of Korea. Sample data from NHIS takes 45 days to receive (deliberation [25 days], payment of fee [15 days], and receipt of data [5 days]). Applicants must submit a research plan and must have obtained IRB approval in advance. The application for HIRA data is similar to that of NHIS; data provision is based on various conditions e.g.: fee must be paid, applicants must detail their affiliated institute, level of education	None identified

Data Protection – Health [Patient consent & exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
information on disease and health care resource utilisation. Data must be de-identified before use, through: (1) preliminary review, (2) de-identification, (3) propriety assessment, (4) post management. Additionally, NHIS and HIRA have their own internal instructions and guidelines (Operating Regulations on National Health Information Data Provision, and Operation Guideline on Public Data Provision and Use, respectively). Richer clinical and cost information is collected through electronic medical records. However, the quality of data collection varies.	evaluation group and assessment procedure). Patient data can be linked between datasets only within the same database i.e. datasets within HIRA's national sample data can be linked together with a common variable (e.g. key sequence number of claim), or NHIS NSC data could be linked across different NHIS datasets by linking via a patient's alternative serial-number ID. However, data cannot be linked between NHIS and HIRA databases (nor with electronic medical records at provider level).	and research title, and must have access to statistical packages. Both NHIS and HIRA data users (once approved) can visit 'big data centres' and operate their analyses on site, which is a requirement for customised data sets (for HIRA customised data, the user may also connect to the analysis server on his/her computer after authentication via remote access system of HIRA). HIRA data users must submit analysis outputs to the director of HIRA within 30 days. Whilst HIRA provides data to both public and private sectors, NHIS is more restrictive to researchers in public policy and academia only.	

Figure 8. Heat map of data governance arrangements in South Korea

	Routinely collected /De novo Raw data	**South Korea**
Government as Regulator	✓Data protection legislation (health 'special case')	The general law governing data protection is the Personal Information Protection Act 2011 (PIPA). However it is not clear to what extent health is viewed as a 'special case' with any corresponding provisions for utilising identifiable data without patient consent.
	✓Equitable patient selection and the protection of vulnerable subjects	
Data subjects: Patients	✓Patient consent	
	✓ Facilitative opt-in / opt-out consent models for research	
Data Collectors	✓Unique patient identifiers (UPIs)	In Korea, everyone has a unique resident registration number (identification number). To ensure the anonymity of a patient, the patient's resident registration number is de-identified, and replaced with an alternative serial-number ID (which is one of the de-identification methods)
	✓Patient information	
	✓Data quality assurance	NHIS and HIRA under obligation to maintain up to date and accurate records under *Personal Information Protection Act 2011*
	✓Data ownership: responsibility for data?	Unclear role for patients in the permissions / management of their data
Data Users	✓Approval of data collection activities to be based on intended use	Researchers must have approval from Institutional Review Board (IRB), which takes into account intended use
	✓ Clear and transparent criteria for de novo data projects	No clear framework for the collection of de novo data
	Cleaning and managing data	
Government as Regulator	✓Data management: Recognised data stewardship entities	The collectors of the data (NHIS and HIRA) have responsibility for its management and access
Data Controllers	✓Process for de-identification	Four-step process, encompassing: preliminary review, de-identification, propriety assessment and post management
	✓Security arrangements: 'Privacy Enhancing Techniques and Procedures' (PETs)	
	✓Training of staff	
	✓Specified arrangements for how long data are kept	
	Linkage and aggregation	
Government as Regulator	✓Privacy rules	Clear process of assessment of privacy risk
	✓Develop a clear set of nationally agreed and implemented standard rules to optimise interoperability of health record systems	Datasets cannot be linked between databases (e.g. cannot link HIRA and NHIS data)

	Routinely collected /De novo Raw data	**South Korea**
Data Controllers	✓Unique patient identifiers	Alternative serial-number ID, which is a de-identified version of the resident registration number
	✓Pseudonymisation	Process for pseudonymisation exists where it is necessary to keep identifying information
	✓Preparation for sharing	Processes outlined by the data controllers
	Access / use of data	
Government as Regulator	✓Managing re-identification risk	
	✓Criteria for different uses (& different users)	More abundant data are provided to users with academic/public policy research purposes, whereas limited data are provided to users in private sectors, such as pharmaceutical companies or individual entities
Data Controllers	✓Approval panels	Review committee of NHIS, HIRA
	✓Confidentiality and data use agreements	Personal Information Protection Act 2011 (PIPA)
	✓Balancing benefits of linkage for research with risk for re-identification	
Data users	✓Audit / Service evaluation and quality monitoring	
	✓Degree of access, level of data, and mode of access	Processes clearly set out, but the access to data is limited to academic purpose
	✓Cost of access	Cost of data access clearly outlined
	✓Appropriate experience/qualifications, and funding to conduct research	

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement. Source: Based on authors' interpretation.

4. CONCLUDING REMARKS

As argued in Cole et al. (2015), the evidence that is used to support decision-making in health care is becoming increasingly diverse, reflecting the increased complexity of the regulatory and reimbursement processes. Increasingly, the importance of understanding the impact of health care interventions in real-world settings is being recognised. In this report, we describe the process by which RWD (the raw data) is transformed into RWE (the insight) in South Korea, and assess the rules and roles for information governance along this process, comparing them to the "ideal framework" proposed in Cole et al. (2015). It seems South Korea compares relatively well with that "ideal framework", although there is room for improvement – in terms of clarifying patient consent for seeking permissions/management of their data, and improving linkages across datasets.

REFERENCES

Cole, A. M., Garrison, L. P., Mestre-Ferrandiz, J. and Towse, A., 2015. *Data Governance Arrangements for Real-World Evidence*. London: OHE Consulting Report. Office of Health Economics. Available: https://www.ohe.org/publications/data-governance-arrangements-real-world-evidence. [last accessed: 21 July 2017]

Kwon, S., Lee, T. and Kim, C., 2015. *Republic of Korea health system review*. World Health Organization. Manila: WHO Regional Office for the Western Pacific Volume 5, no.4, Available: http://iris.wpro.who.int/handle/10665.1/11358. [last accessed: 21 July 2017]

Lee, J., Lee, J. S., Park, S.-H., Shin, S. A. and Kim, K., 2016. Cohort Profile: The National Health Insurance Service–National Sample Cohort (NHIS-NSC), South Korea. *International journal of epidemiology,* dyv319.

APPENDIX 1 PRO-FORMA

1. **Brief overview** of the health system and collection / management of patient data. Specify the key data sources available for RWD.

2. **Core legislation and governance arrangements for the collection and/or use of patient data**
   a. **Routinely collected patient data.**
      Core legislation governing the collection / use of routinely collected patient data. Review and summarise key documentation outlining principles of governance and data protection.
   b. **Collecting de novo patient data.**
      Governance arrangements for research to collect new data. Review and summarise key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

3. **Data linking.** To what extent can patient data be linked across datasets? What are the organisations involved, and what are the core governing principles under which they operate?

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

5. **Data use.** What, if any, are the rules governing the use of RWD? [To cover contract arrangements between data suppliers and recipients, rules around use for HTA, etc.]

6. **Governance ideals and changes to the environment.** Summarise any key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.