Strategy Considerations in Genome Cohort Construction in Korea

Joohoon Sung, Sung-il Cho1)

Department of Preventive Medicine, Kangwon National University College of Medicine, Department of Epidemiology, Seoul National University School of Public Health1)

Focusing on complex diseases of public health significance, strategic issues regarding the ongoing Korean Genome Cohort were reviewed: target size and diseases, measurements, study design issues, and follow-up strategy of the cohort. Considering the epidemiologic characteristics of Korean population as well as strengths and drawbacks of current research environment, we tried to tailor the experience of other existing cohorts into proposals for this Korean study. Currently 100,000 individuals have been participating the new Genome Cohort in Korea. Target size of de novo collection is recommended to be set as between 300,000 to 500,000. This target size would allow acceptable power to detect genetic and environmental factors of moderate effect size and possible interactions between them. Family units and/or special subgroups are recommended to parallel main body of adult individuals to increase the overall efficiency of the study.

Given that response rate to the conventional re-contact method may not be satisfactory, successful follow-up is the main key to the achievement of the Korean Genome Cohort. Access to the central database such as National Health Insurance data can provide enormous potential for near-complete case detection. Efforts to build consensus amongst scientists from broad fields and stakeholders are crucial to unleash the centralized database as well as to refine the commitment of this national project.


Key words: Cohort studies; Human genome; Genetic predisposition to disease; National Health Insurance, Republic of Korea; Follow-up studies; Biological specimen banks
Table 1. Estimated disease prevalence (per 1,000 persons per year), incidence (per 100,000 persons per year) and mortality (per 100,000 persons per year) rates of major disease groups according to the disease classification of global burden of disease study

<table>
<thead>
<tr>
<th>Disease group</th>
<th>Prevalence</th>
<th>Incidence</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td><strong>Stomach cancer</strong></td>
<td>1.2</td>
<td>0.7</td>
<td>48.1</td>
</tr>
<tr>
<td><strong>Colon and rectum cancers</strong></td>
<td>0.6</td>
<td>0.5</td>
<td>20.9</td>
</tr>
<tr>
<td><strong>Liver cancer</strong></td>
<td>0.8</td>
<td>0.2</td>
<td>34.9</td>
</tr>
<tr>
<td><strong>Pancreatic cancer</strong></td>
<td>0.1</td>
<td>0.1</td>
<td>61</td>
</tr>
<tr>
<td><strong>Trachea, bronchus and lung cancers</strong></td>
<td>0.6</td>
<td>0.2</td>
<td>31.7</td>
</tr>
<tr>
<td><strong>Breast cancer</strong></td>
<td>0.0</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Cervical uteri cancer</strong></td>
<td>0.1</td>
<td>0.1</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>Prostate cancer</strong></td>
<td>0.1</td>
<td>0.1</td>
<td>5.2</td>
</tr>
<tr>
<td><strong>Lymphoma and multiple myeloma</strong></td>
<td>0.01</td>
<td>0.1</td>
<td>3.9</td>
</tr>
<tr>
<td><strong>Hodgkin disease</strong></td>
<td>0.0</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Alcohol abuse</strong></td>
<td>0.0</td>
<td>0.1</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Schizophrenia</strong></td>
<td>2.2</td>
<td>1.9</td>
<td>48.9</td>
</tr>
<tr>
<td><strong>Unipolar major depression</strong></td>
<td>1.3</td>
<td>2.8</td>
<td>49.3</td>
</tr>
<tr>
<td><strong>Bipolar disorder</strong></td>
<td>0.4</td>
<td>0.4</td>
<td>10.7</td>
</tr>
<tr>
<td><strong>Dementia and degenerative CNS disorder</strong></td>
<td>0.3</td>
<td>0.5</td>
<td>13.0</td>
</tr>
<tr>
<td><strong>Epilepsy</strong></td>
<td>2.4</td>
<td>1.8</td>
<td>57.9</td>
</tr>
<tr>
<td><strong>Parkinson disease</strong></td>
<td>0.3</td>
<td>0.4</td>
<td>6.8</td>
</tr>
<tr>
<td><strong>Glaucoma</strong></td>
<td>1.1</td>
<td>1.2</td>
<td>25.6</td>
</tr>
<tr>
<td><strong>Cataracts</strong></td>
<td>1.4</td>
<td>2.4</td>
<td>53.9</td>
</tr>
<tr>
<td><strong>Diabetes mellitus</strong></td>
<td>1.5</td>
<td>2.2</td>
<td>181.3</td>
</tr>
<tr>
<td><strong>Ischemic heart disease</strong></td>
<td>2.3</td>
<td>1.7</td>
<td>63.6</td>
</tr>
<tr>
<td><strong>CVA (cerebrovascular attack)</strong></td>
<td>3.6</td>
<td>3.5</td>
<td>125.7</td>
</tr>
<tr>
<td><strong>COPD (chronic obstructive pulmonary disease)</strong></td>
<td>7.5</td>
<td>7.1</td>
<td>215.8</td>
</tr>
<tr>
<td><strong>Anemia</strong></td>
<td>10.9</td>
<td>9.3</td>
<td>272.0</td>
</tr>
<tr>
<td><strong>Pepsic ulcer disease</strong></td>
<td>16.5</td>
<td>19.5</td>
<td>406.0</td>
</tr>
<tr>
<td><strong>Colitis of the liver</strong></td>
<td>3.6</td>
<td>0.8</td>
<td>102.6</td>
</tr>
<tr>
<td><strong>Reumatic arthritis</strong></td>
<td>0.7</td>
<td>2.8</td>
<td>18.4</td>
</tr>
<tr>
<td><strong>Osteoarthritis</strong></td>
<td>2.8</td>
<td>11.5</td>
<td>74.2</td>
</tr>
<tr>
<td><strong>Congenital Anomalies</strong></td>
<td>0.5</td>
<td>0.3</td>
<td>28.3</td>
</tr>
</tbody>
</table>
가 기여할 수 있는 장점은 무엇이며 이를 어떻게 극대화할 수 있을 것인지에 대한 문제의식을 중심으로 기술하고자 하였다.

본론

1. 유전체 코호트의 주요 대상질환과 필요요코호트의 주요 대상질환과 필요요


이러한 주요 질환의 발생률, 유병률, 사망률 등 주요 변도표표는 Table 1에 제시하였다. 질병의 변동은 약 100만 명의 표본을 우리나라 전부 인구에 대해서 성별, 연령별, 지역별 대표성을 갖도록 무작위 추출한 후, 이들의 질병발생 및 사망양상을 (1998년 ~ 2002년 사이) 의료보험 정책자료와 활동자료, 사망원자료 등 의 가능한 자료들을 토대로 1차로 검색하였다. 1차로 검색한 결과들은 환자 기준을 작성하였고 대상 환자로 포함하였다. 예를 들어, 뇌졸중의 경우는 입원의 기록이 있고 입원기간이 2일 이상이면서 2회 이상 동일한 진단이 4년 안에 있을 경우를 뇌졸중으로 인정하였다. 이렇게 산출된 질환의 변도표표들은 다시 2,000명 정도의 표본에 대한 실제의 무기록 조사를 통해 변동을 보정하여 최종적으로 표 1의 결과를 얻었다. 우리나라의 주요 질병부담을 구성하고 있는 질환들은 10년 만의 발생양상이 30명에서 200명 사이인 경우가 대부분임을 알 수 있다 [12]. 이러한 질환들이 모두 유전적, 환경적 요인들에 의해서 유전적 진행, 유전자의 상호작용을 통해 질환의 위험성이 결정된다고 하면서, 몇 가지의 가정을 통해서 필요한 코호트의 규모를 예상해 볼 수 있다. UK Biobank는 동일한 문제의식을 가지고 약 50만 명 규모의 연구집단을 구성하여 약 20년간 추적관찰을 하는 것이 필요할 것으로 주장하였다 [7]. 또한, 미국의 국립건강 유전체연구소 (National Human Genome Research Institute)에서의 위험요인을 유전적 위험요인, 환경적 위험요인, 유전-환경의 상호작용, 유전적 건강의 상호작용 등으로 나누어 평가하였으며, 5년간의 추적조사를 통해서 파악할 수 있는 요인들은 10만명 당 50명 정도의 발생물을 보이는 질환으로 대상

Table 2. Estimated number of cases and power to detect the smallest effect size (as relative risk, RR), according to the different target cohort size, by the incidence rates, etiologic factors (genetic or environmental), and follow-up duration (5 year and 10 year follow-up), adapted from Manolio et al

<table>
<thead>
<tr>
<th>IR</th>
<th>Disease examples</th>
<th># of cases</th>
<th>Detectable Effect Size (RR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>E</td>
<td>GxE</td>
<td></td>
</tr>
<tr>
<td>Follow-up duration (years)</td>
<td>Follow-up duration (years)</td>
<td>Follow-up duration (years)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>10 Parkinson disease, schizophrenia</td>
<td>91</td>
<td>201</td>
<td>40</td>
</tr>
<tr>
<td>50 Colorectal cancer, renal failure</td>
<td>456</td>
<td>1,008</td>
<td>19</td>
</tr>
<tr>
<td>100 Breast cancer, hip fracture</td>
<td>912</td>
<td>2,016</td>
<td>1.5</td>
</tr>
<tr>
<td>300 Diabetes, stroke, heart failure</td>
<td>1,820</td>
<td>4,022</td>
<td>1.4</td>
</tr>
<tr>
<td>500 Myocardial infarction, all cancers</td>
<td>4,524</td>
<td>9,998</td>
<td>1.3</td>
</tr>
<tr>
<td>3,000 Cardiovascular, hypertension</td>
<td>25,858</td>
<td>57,146</td>
<td>1.1</td>
</tr>
</tbody>
</table>

IR - incidence rate per 100,000 persons per year * G-genetic effect, E-environmental effect, GxE-gene by environmental interaction effect * aging effect was also considered in calculating the number of cases. Bold figure denotes for acceptable level of detection power, here effect size of 1.3, 1.5 and 2.0 were chosen for genetic, environmental and gene-environmental interaction, respectively.
Table 3. Type of recruited families and sampling efficiency for typical genetic analyses

<table>
<thead>
<tr>
<th>Family unit</th>
<th>Contributions to each analysis</th>
<th>Sampling efficiency (per person)</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pedigree examples</td>
<td>Classic Linkage</td>
<td>IBD Linkage</td>
</tr>
<tr>
<td>Family History</td>
<td>?</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Affected relative pairs (sibs or other)</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Trio</td>
<td>0</td>
<td>&lt;2</td>
<td>1</td>
</tr>
<tr>
<td>Nuclear Pedigree (n=4)</td>
<td>&lt;4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Large Pedigree (n=7)</td>
<td>10</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Large pedigree with twin (n=7)</td>
<td>8</td>
<td>19</td>
<td>4&lt;</td>
</tr>
<tr>
<td>Incomplete large pedigree (n=4)</td>
<td>&lt;8</td>
<td>6</td>
<td>&lt;4</td>
</tr>
</tbody>
</table>

1. Actual power for IBD-based linkage drops more because parent-offspring pair does not have IBD distribution (i.e., fixed to 0.5)
2. Incomplete pedigree may result in severe loss of power, especially when the marker information is not high.

### 2. 기존에 구축된 코호트와 새로운 유전체 코호트

현재 우리나라에서 비슷한 수준의 환경적 요인에 대한 조사가 수행되고, 유전체 및 생체시료들이 확보되고 있는 코호트는 다각관 암코호트 등이 있으며, 확인된 코호트의 구성을 현재 이어 9만 명에 이르고 있고 장기적으로는 약 20만 명 수준에 도달할 것으로 기대된다 [14]. 기존의 코호트와 통합하여 유전체 코호트가 구성될 것이며 혹은 새로운 유전체 코호트만으로 필요한 표본수를 재워나갈 것인지는 특히 미국에서 큰 논쟁이 진행되고 있다 [15, 16]. 미국에서는 국가가 주도하는 유전체 코호트의 출발이 영국, 일본 등과 같이 우리나라에 비해서도 늦었으며, 기존에
구축되어 온 대규모의 코호트가 이미 집중되어 있는 상황이기 때문에 각각의 국가적인 투자가 시급한 편에 대해서 논란이 이루어지는 것은 당연하다. 하지만 우리나라의 경우 기존에 구축된 대규모 코호트가 거의 없고, 새로운 구축되어 가고 있는 유전체 코호트가 이미 성공적으로 진행되고 있다는 점, 또한, 기존의 코호트들도 비슷한 문제의식과 프로토콜을 가지고 충만한 경우가 많기 때문에 상당 부분의 내용들이 호환될 수 있는 가능성이 있다는 점 등은 미국과 다른 상황이라고 할 수 있다 [17]. 즉, 우리나라의 상황에서는 현재와 같이 유전체 코호트가 중심이 되어 독자적으로 30만 명 정도 혹은 그 이상의 코호트를 확보하는 것이 필 요하다고 판단된다. 이것은 다가간 양 코호트, 원전 코호트 등의 다양한 코호틀들이 호환가능한 프로토콜을 가지고 진행될 수 있다고 할 때 현실적으로는 약 50만 명 규모의 코호트가 구축된 것과 같은 효과라고 할 수 있다.

3. 코호트의 모집과 참가자에서 수집되는 정보

유전체 코호트는 프로토콜의 통일성 제고를 위한 별도의 연구과제가 수행되는 등 조사구도와 조사내용, 수집되는 정보의 수준을 제고하고 표준화하는 노력이 진행되고 있다. 표준화와 통일성 제고는 코호트의 성과를 좌우하는 매우 중요 한 핵심적인 과제이다. 특히 코호트 연구의 특성상 환경요인의 측정이 확립되어 정확하게 그리고 각 단위 간에 통일적으 로 이루어 질 수 있도록 많은 노력을 기울여야 할 점을 강조할 필요가 있다. 즉, 환경요인을 질병발생 이전의 시점에서 정확하게 평가하지 못한다면, 그의 투자가 투자 도해 코호트를 구축하여 하는 의의가 반감되기 때문이다. 유전요인은 일생을 통해 여러 변화가 없고 현실에 법적 효과적인 환경-대조군 연구에서도 독립적으로 평가될 수 있음을 명확히 한다.

수집되는 정보의 또 다른 축인 생체운영을 위한 경우 대규모의 연구에서 적합한 체계적인 biobank의 구축과 운영, 관리가 시급히 요구된다. 이러한 biobank의 존재는 향후 유전체학, 단백체학, 프로토콜 등에 대해 비례의 과학기술의 발전에 따라 그 유도를 지속적으로 개발될 수 있다 [16,15]. 따라서, 양질의 생체서로를 확보하기 위한 노력에 극히 중요하고 현 시점이 마련된 표준 운영지침(Standard Operating Procedure, SOP)을 확대 강화하여 모든 코호트 구축의 단위 기관들에게 엄격하게 적용하여야 할 것이다.

4. 코호트의 투입적인 목표달성을 위한 특수집단의 선정

코호트의 규모에 관한 문제에서 반드시 짐을 남겨야 할 문제는 유전적 요인에 관한 가정이 너무 단순하다는 점이다. Manolio 등은 추정 "다변수 유전요인에 의한 다변등질집단"(common disease common variant, CDCV)의 개념을 전제로 하여, 10% 정도의 변동을 가진 단일한 유전요인이 질환의 발생 위험을 높이는 것을 기본 전제로 하고 있다 [4]. 그러나, CDCV 가설은 많은 사람들의 기대에도 불구하고, 아직 변호 성립적적으로 검증된 바가 없으며, 현재 덜 낮은 변동을 가진 유전요인들의 집단이 다변등질의 유전적 표현형의 가능성이 얼마든지 있다. 또한, 매우 유전 요인을 가진 사람에서에서 집단이 발현되는 경우를 " genomewide"의 개념이 유전적 표현형의 가능성을 나타내고 있다. 즉, 해당 대규모유전자가 하나만 있었다면 집단의 위험도가 증가하는 가정으로 다를 수 없으나, 가산적인 공동 유전형(co-dominant), 즉 시행 선정유전자가 갖는 경우에 위험도가 멀어질 수 있는 경우, 혹은 열성이(recessive)의 형태, 즉, 두 가지의 유전형을 가질 때만 비로소 질환발생의 위험이 증가하는 형태 등이 가능성 있는 가치들이 뒷받침되며, 가설적 시장의 측정이 저선호도에 부합하다. 이러한 가설은 단편적에 있어 약간 주관적인 것이라는 비판을 면할 수 없다. 유전체학적 연구는 유전체학적 발현요인의 체계적 성 투영연관성을 입으로 발생하는 산발적인 발생 환경(sporadic cases) 등의 비중이 매우 높도록 예상되며, 이러한 경우는 활현 더 많은 환경이 있어야만 동일한 위험도를 가유한 유전 요인을 규명할 수 있다는 점도 강조되어야 한다. 이런 다양한 유전적 기전들이 있다는 것이 가족형질환에서 예외 없는 복합질 화자의 정의가 있기 때문에, 유전적 요 인에 대한 접근은 일반인구에 대한 접근 방식에 더할 뿐만으로는 해결되지 어려운 한계가 있다. 이를 극복하기 위해서는 일반인구 단위의 모집전략과 병행하여 가족단위에 대한 모 집단을 수집할 필요가 있다. 예를 들어, 가족단위의 모집이 진행될 경우, 유전자 연결분석(linkage analysis)을 통해 genemapping 연구를 수행할 수 있기 때문에, 후 보 유전자의 액체생 선관후의 위치를 제시해 줄 수 있게 된다 [18]. 향후, 전장 유전체(genome-wide)에서도 수집반영의 SNP를 검색하는 방법이 표준적인 분석방법이 된다고 가정하면, 각 연구를 통해 제시되는 후보 유전자들의 존재는 최대의 난제 중 하나인 다중비교의 문제를 해결할 수 있는 가장 좋은 도구가 되어야 할 것이다. 또한, 환경적 요인에서 생물기법의 중요성을 가지는 특정 시기의 노출은 해당 연구군이나 해당 인구집단에 대한 조사를 통해서 보완될 필요가 있다. 예를 들어, 대 대 난 환경은 평생 동안의 건강수준을 유 유할 수 있는 중요한 기간으로 평가되고 있지만, 생전은 물론, 학습기의 정도의 소망이 되어도 정확한 측정이 좀 더 불가능한 요인이다 [19]. 생계정책의 특수집단 단 이외에도, 도시와 농촌, 다양한 사회경제적 구조의 계층 등에 대한 고려를 통해서 환경적 요인의 변화가 충분히 확인될 수 있고, 결과의 일반화가 가능하도록 연구내 용의 확보가 고려될 필요가 있다. 우리나라의 다민족적 집단이라는 기존의 고정관념에서 벗어난 현실을 보는 데, 극 중장하는 국제경쟁으로 이미 상당수의 외국 출신 한국인이나 외국 출신 한국인이 조사되고 있는 현실이다. 이러한 경우는 연구와 성별에 대한 고려들을이나 다양한 민족적인 배경에 대한 고려가 모든 연구를 정책수립의 필수적인 존재가 되고 있다. 코호트가 최소한 10 년에서 20년에 대한 연구성과가 나타나는 것이라고 할 때, 현재 증가하고 있는 외국 인이민들은 우리 사회의 한 구성원으
로 자라 색이 확실하며, 역시 코호트 에 포함되므로 혈특수질환을 알 것이다.

5. 가족단위 연구의 병행을 위한 고려 사항
가족단위의 연구는 복합질환의 연구에
서도 몇 가지의 고유한 장점을 가지고 있
으며, 이 때문에 미국의 유전체 코호트 구
축을 위해서도 중요한 연구과정의 하나로
거론되고 있다. 따라서 가족 연구의 고려
사항을 별도의 주제로 간략히 다루어보고
자 한다. 가족 연구가 기여할 수 있는 장점
은 다음과 같다. 1) 가족적인 집단(familial aggregation)을 보이는 환자들, sporadic case 가 아닌, 유전체 연구의 목표인 유전적인
감수성의 공동체에 의한 확률이 더 높다. 2) 가족적인 집단을 보이는 환자들 중 일부는 유전적인 요인이 작용할 확률이 현행관계와 없는 환자들 사이에서 보다 높게 놓아
서 보다 큰 견정력을 가진다. 예를 들어, 한
명의 유방암 환자에서는 A 및 B 라는 유전
요인과 과장적 석, 경기치의 이상 호르
몬 노출 등이 질환발생의 원인이 되었고,
또 다른 유방암 환자에서는 C 및 D 라는 유전
요인과 흡연, 환경성 방사물질 및 질
환발생의 원인이 되었다고 가정하면,
일반인구들 대상으로 한 연구에서는 새로
운 환자가 추가되어 A 및 B 라는 유전요
인과 C 및 D 라는 유전요인은 기존의 모든
분석에서 상쇄되어 위험요인으로 받아
기 어렵게 된다. 3) 가족적 형태의 연구에서
는 기존에 알려진 지식과 전문 무관하
게 유전자가 위치하고 있는 유전자
화를 제시하는 gene mapping 연구를 수행
할 수 있어서, 연구결과 연구의 상승요
과를 할 수 있게 한다. 분석방법을 최근 긍
격한 진화를 보이고 있으며, 복합질환을
위한 방법론들이 속속 개발되고 있다.
(2021) 4) 일반인구의 전장유전체 분석에
통상 50만 ~ 60만 개정도의 유전형 분석이
필요한 것과 달리, 가족연구는 400 ~
1,000개의 유전형 분석만으로도 충분히 결
과가 나올 수 있어서 비용효과가 있다.
반면, 가족 단위연구는 몇 가지 단점을 가지고 있어서 1) 일반인구의 개인단위 모
질에 비해 가족 전체를 포괄시키기 위
한 노력이 훨씬 더 힘들다. 2) 가족은 가족
의 크기에 따라서 견정력에 큰 차이를 가
저오게 되어, 예를 들어 100 명의 가족을
모집한 경우라도, 3일간의 조건으로 30
여 가족을 모집한 경우(60 명) 보다 8인 가
족으로 11 가족을 모집한 경우(88 명)의 견
정력이 10배 이상 커지게 된다. 가족의 크
기별로 1) 고전적인 linkage analysis, 2) IBD (identity by descent)를 기반으로 한 penetra-
nce model-free linkage analysis, 3) TDT (transmission disequilibrium test) 분석 등
에서 실제 기여하는 표본수 및 표본효율(한
사람당의 견정력)을 Table 3에 제시하였다.
가족 단위의 코호트는 약 1년 5천 ~ 2만
명 정도의 규모로 목표로한 경우 Table 2
와 동일한 가정 하에서 대부분의 주요 질
환들에 대한 분석이 가능할 것으로 추정
되며, 약 4천 5백만의 인원을 기준의 생존
이 가족 연구에 (2022), 약 3 만 명 정도를
영유아 및 소아를 대상으로 하는 성장발
달 코호트에서, 또한 나머지 지역사회
단위 코호트의 가족 하부단위와 일반 코
호트에서 가족적이 있는 환자들의 가족
생의 모니터링은 동일하게 할 수 있을 것으로 전망된다.

6. 국가 건강정보를 활용한 추구조사
의 전략
기존의 코호트는 추구조사에 대한 명확
한 지시와 전략을 가지고 있지 못한다. 약
성중양에 대해서는 암등록을 활용할 수
하지만, 다른 질환의 발생을 확인하기 위
해서는 일정이 제한적이다. 실제로는
현재는 추구조사에 대한 응답률이 어느
정도인지에 대한 기본 자료도 확보되어
있지 못하다. 예를 들어, 만일 2년 주기로
70% 만 추구조사가 되더라도 가정하면, 이론
상으로는 5년후에는 약 41, 10년 후에는
16.8%의 대상자 추구조사가 되고 있다.
이는 제한적이다. 한 사람 한 사람을 연
구에 참여하기 위해서 흔히 차원과 노
력을 강화한 때, 추구조사의 방법론을 확
보하는 것은 코호트의 성공과 좌우하는
핵심문제라고 생각된다.
우리나라는 1980년대 후반부터 20여 년
에 이르는 건강정보가 국가의 주도로 건
강보험에 의해서 관리되어 오고 있다. 영
국의 UK Biobank가 영국의 국가건강정보
제도(National Health Service, NHS)를 통해
서 질환의 발생을 거의 완전하게 파악할
수 있다는 자산감을 바탕으로 추진되고 있고 [4, Iceland의 deCODE biobank가 역시
국민들의 건강질병 정보가 집계되어 있
다면 부과세를 창출하고 있는 사례 [9]
들은 우리나라의 국가 자원이 활용되어야 할 필요성을 유래해 주고 있다. 물론, 우
리나라의 건강정보자료가 실제 추구조사
에 활용되기 위해서는 제도적인 차원의
문제나 개인의 동의를 구하는 응용적인
자료의 문제뿐만 아니라, 자료를 정확하
게 활용할 수 있기 위한 방법론의 개발도
필요하다. 그러나 이러한 과학적인 차원
의 문제들은 활용 가능성이 열린다면 큰
문제 없이 해결될 수 있을 만큼 여러 연구
자들에 의해서 활용방안이 개발되고 경
청이 촉발되어 왔다 [23-25].
현재 건강정보자료 등이 개인정보를 보
호하는 장치가 부족하기 때문에, 사용을
제한하기 위한 법률조치들이 제출되고 있
으나, 개인정보의 철���한 보호와 광범한
공익 활동계획을 수립하는 것은 동일의
양면이다. 즉, 활용을 전제로 한 개인정보
가 보호될 만큼의 기술적이고 이미 확보
되어야겠지만, 지금처럼 활용이 제한될 때
는 개인정보 보호에 대한 중요성도 갑작
되고 개별정보 유출사고가 발생될 수도
있는 것이다. 유전체 코호트는 물론, 국가
의 보건의료 관련 연구사업에 있어서, 1) 동의
를 구한 사람들의 대한 질환발생의 확인,
2) 개인 식별이 불가능하도록 암호화 된
공익목적의 데이터베이스가 구축되는 일
(예, 약물부작용의 모니터링 혹은 질병 발
생의 위험요인 평가) 등은 시급한 과제이
다. 단지 유전체 코호트 연구를 위해서 뿐
만 아니라 세계적으로도 무한한 활용가
치를 인정받고 있는 우리나라의 소중한
지식 자료로어서, 필요한 공익활용을 위해
주요약도 모아지고 활용방안이 추구되
어야 할 것이다.

결 론
우리나라의 유전체 코호트는 우리가 가