

2018

English Research Paper – 26

Improvement and discovery of policy through utilization of Gender Big Data(I)

Youkyung Moon

Kitaek Jeon

Hojoong Bae

Keuntae Kim



Korean Women's Development Institute

**Improvement and discovery of policy
through utilization of Gender Big Data(I)**



©2018

Korean Women's Development Institute

225 Jinheung-ro Eunpyeong-gu

Seoul 03367

Korea

www.kwdi.re.kr

Contents

I. Introduction	1
1. Purpose of research	1
2. Content of Research and Methodology	2
II. Background of Big Data Research on Women and Family	8
1. Concept and Categories of Big Data for Women and Family	8
2. Women's Policy Relations in Big Data Policy	10
3. International Organizations and Overseas Cases	11
III. Availability of Public data on Women and Family Policy	13
1. Background and Purpose of Research	13
2. Status of Public Data for Women and Family	14
3. Analysis of the Availability of Big Data for Women and Family	15
4. A Pilot Study of Big Data with Women's Family Public Data: Analysis of the Demand and Supply of Child Rearing and Care	17

IV. A Study on the Public Opinion of a Childbirth Policy Using Data Science	19
1. Online Big Data Analytics	19
2. FGI Analysis of the People Affected by Policy	21
3. An Analysis of Major Presidential Candidates' Campaign Pledges	23
4. Conclusion	23
V. Analysis of the Survival Rates of Self-Employed Businesswomen Using Credit Card Big Data	25
1. Purpose of Research	25
2. Building Big Data Materials	25
3. Analysis Results	26
VI. An Analysis on the Status of Re-employment of Women with Career Interruption Using Public Big Data	29
1. Purpose of Research	29
2. Building Big Data Materials	30
3. Analysis Results	30
VII. Policy Proposal	32
1. Policy Suggestions by Subject of Big Data Analysis	32
2. Policy Proposals for Big Data and Production	37

Table

⟨Table 1⟩	List of Task and implementing institutions	4
⟨Table 2⟩	Big Data Research Methodology and Gender Forum ...	4
⟨Table 3⟩	Big Data Symposium on Women and the Family	5
⟨Table 4⟩	Notes on Attended Big Data Seminar and Advisory Council	6
⟨Table 5⟩	Expected Topics for Gender Analysis of Big Data Materials	9

Figure

〈Figure 1〉 Research plan and content by year 3

Improvement and discovery of policy through utilization of Gender Big Data(I)

Youkyung Moon

Kitaek Jeon

Hojoong Bae

Keuntae Kim

I . Introduction

1. Purpose of research

- Despite the overall development of policies regarding big data, discussions on how to use this data for gender policy have not been widely activated. The use of big data in the establishment and implementation of women's policies is expected to bring new results and development.
- The current government administration established a five-year plan to actively utilize big data as a means of collecting public opinion. The administration also plans to use big data for effective policy development to implement an intelligent government equipped with advanced technology such as artificial intelligence and big data analysis, as well as to develop innovative financial services such as big data fin-tech. In addition, the government plans to utilize big data to address problems in civil life, such as early detection of children at risk and public against natural disasters and accidents.

- The purpose of this study are first to explore how big data analysis and related government policies can be linked to gender policies. Secondly is to attempt exemplary big data analysis through a trial conducted on major issues related to women and families. Third is to prepare a suitable environment for the production of big data on women and family issues by combining diverse data from various departments and agencies related to gender and family state policies.

2. Content of Research and Methodology

A. Research content

- This study is conducted as a three-year plan and this report outlines the findings of the first year of research. The three years of research can be divided into two broad categories. The first category is the creation of a workable environment for the utilization of big data on women and family issues. Due to the nature of big data research, it is necessary to collect raw materials and then convert or process that material into analytical forms. The idea of big data analytics may actually fail or put excessive resources into the data collection process. In order to improve the availability and accessibility of big data, we will conduct basic research on big data collection and processing.
- The second category of research is trial analysis using big data and the subsequent derivation of policy improvement. To the extent that the data can be collected, the trial program attempts to leverage big data resources to generate substantial policy improvements.

	Building environment for utilizing gender/family big data	Pilot analysis by topics and suggestions for policy improvement
1 st Year (2017)	Utilizing administrative data of gender/family policies as big data.	SNS Analysis on low birth rate Big data analysis on women employment
2 nd Year (2018)	Potential of converting public sector big data for gender/family	Big data analysis on senior women health Big data analysis on women and safety
3 rd Year (2019)	Implementation, sharing and networking big data platforms	Big data analysis on women's finance, credit, tax Big data analysis on discrimination in labor market

〈Figure 1〉 Research plan and content by year

- During the first year of research, the program’s focus is on the potential for utilizing administrative and statistical data generated in the implementation of gender and family policies.
 - The status of production and deployment of big data related to the women and family policies, such as gender equality, workforce development and human rights protection, was analyzed and the possibility of utilization was reviewed.
- The following tasks will be conducted for trial analysis and policy improvement
 - Public opinion analysis on low birthrate using SNS data
 - Analysis of the survival rate of self-employed women using credit card big data
 - Analysis of the reemployment of women with interrupted careers using administrative big data.

B. Research Methodology

- Collecting domestic and international literature and data
 - Collection and review of research literature and policy data related to big data and public policy
 - Collecting and reviewing research literature and policy data related to women and big data
- Collection and review of big data cases related to women and families in the private and public sectors
 - Collection of cases from international organizations and major developed countries
 - Collection and analysis of raw materials of domestic big data
- Conducting collaborative studies and commissioning tasks to professional organizations to interact with various majors and sectors.

〈Table 1〉 List of Task and implementing institutions

Task 1 : Analysis of public opinion regarding birth policy using data science (by Arspraxia)
 Task 2 : Potential for using public data on Women and family policy and trial Analysis (by the Korea Institute for Health and Social Affairs)

- Holding two multi-disciplinary hybrid forums to raise women researchers' interest in big data.

〈Table 2〉 Big Data Research Methodology and Gender Forum

1st Women and Big Data Forum : Big Data Research Methodology and Gender Issues
 Date and Location : 2017.6.22. (Thurs) 15:00 – Large Conference Room
 Presentation 1: Theory and Application of semantic network analysis method
 Presentation 2: Public Decision Making Using Shinhan Card Big Data

2nd Women and Big Data Forum : Big Data Research Methodology and Gender Issues

Date and Location : 2017.11.7. (Tues) 15:00 – Large Conference Room on 3rd Floor

Presentation 1: Status and Utilization of National Statistical Office’s Big Data

Presentation 2: Public Health Big Data Deployment : Current Status and Future Prospects

○ Holding ‘Big Data Symposium on Women and Families’

- Intending to examine the current status of women and family big data policies and to present a vision for the establishment of big data utilization for women and their families with an introduction of foreign examples

〈Table 3〉 Big Data Symposium on Women and the Family

Date	Title	Organizer
3.14/ 3.29	Conferences of Experts on Promoting Big Data Cooperative Research	National Research Council for Economics, Humanities and Social Sciences
3.31	Big Data Seminar	Shinhan Bank Big Data Center
4.10	The First Public Data Organization Training	Korea Telecommunications Promotion Association
4.12	Empirical Seminar on Big Data for Personal Information Specification for the Fourth Industrial Revolution	SK Telecom, Korea Association for ICT Promotion, Big Data Forum
4.20	First Sejong Open Forum: Debate on Big Data	Ministry of Personnel Management
4.20	Seminar on Effective Practices for Future Planning and Policy Support using Big Data	Korea Telecommunications Promotion Association
7.25	Conference on The Fourth Industrial Revolution and Big Data: Big Data, Our Lives, Improving Systems for Personal Information Protection in the Age of Big Data	Seunghee Ryu, National Assembly Member
9.14-15	‘Chosun Ilbo Smart Cloud Show 2017’ Introduction to Matrix society – from artificial intelligence and the Cloud revolution to the virtual currency syndrome	Chosun Ilbo
9.25	The 4th Industrial Revolution Era, the Debate on Cyber Security	Seunghee Ryu, National Assembly Member
11.29	8th Big Data – Statistics Strategy Forum	National Statistics Office

〈Table 4〉 Notes on Attended Big Data Seminar and Advisory Council

프로그램 Program

14:00~14:10	10분	개회식 • 개회사 : 권인숙 (한국여성정책연구원 원장)
1부: 빅데이터를 통해 보는 여성문제		사회 : 김경희(한국여성정책연구원 성인지정책연구실장)
14:15-15:35	80분	<ul style="list-style-type: none"> ◆ 발제 1: 데이터 사이언스를 활용한 출산관련 여론 분석과 정책제안 <ul style="list-style-type: none"> • 발표자: 김도훈 (아르스 프락시아 대표) 김학준 (아르스 프락시아 미디어분석팀장) ◆ 발제 2: 여성 자영업자의 생존율 분석 <ul style="list-style-type: none"> • 발표자: 문유경 (한국여성정책연구원 선임연구위원) 배호중 (한국여성정책연구원 전문연구위원) ◆ 발제 3: 빅데이터를 활용한 여성 일자리 분석: 새일센터를 중심으로 <ul style="list-style-type: none"> • 발표자: 김근태 (고려대학교 공공사회·통일외교학부 교수)
〈토론〉		
<ul style="list-style-type: none"> - 박지훈 (고려대학교 신문방송학과 교수) - 류근관 (서울대학교 경제학과 교수) - 윤세진 (여성가족부 경력단절여성지원과장) 		
10분		휴식 및 다과
2부: 여성과 가족에 대한 빅데이터의 구축		사회 : 주재선(한국여성정책연구원 성별영향평가·통계센터장)
15:45-16:45	60분	<ul style="list-style-type: none"> ◆ 발제 4: 여성가족정책 공공데이터 현황 및 활용가능성 검토 <ul style="list-style-type: none"> • 발표자: 오미애 (한국보건사회연구원 빅데이터연구팀장) ◆ 발제 5: 해외의 여성과 빅데이터 정책 현황 <ul style="list-style-type: none"> • 발표자: 전기택 (한국여성정책연구원 여성고용연구센터장)
〈토론〉		
<ul style="list-style-type: none"> - 김성현 (한국정보화진흥원 K-ICT 빅데이터센터 수석연구위원) - 정용찬 (정보통신정책연구원 미디어통계분석그룹장) 		
16:45-17:00	15분	◆ 질의 & 응답

- Active exchange of opinions by holding step-by-step expert meetings, launching report meetings and interim report meetings for regular and commissioned tasks.
 - Includes academia, government, private research institutes, and businesses.
 - Interaction with various majors such as sociology, statistics, computer science, women's studies, social welfare studies, education engineering, etc.
- Attend major domestic big data forums, training and advisory meetings, etc.
 - Creating consensus on the research results of the big data on women and families
 - Networking with big data workers and researchers
 - Amassing relevant information and cases

II . Background of Big Data Research on Women and Family

1. Concept and Categories of Big Data for Women and Family

- Big data on women and the family is closer in nature to big data ‘for’ women and the family than just big data about women and the family.
 - The boundaries of big data for women and families should include data for research and policy development on gender equality, women’s status improvement and the gender gap, plus big data on the creation and disappearance of families and on family relationship
- The concept of ‘big data for Women and Families’ can be defined as two-fold
 - The broader scope of definition is ‘inclusion of the element of gender awareness in all big data and subsequent reevaluation and rebuilding of existing big data with this perspective’.
 - For example, big data on the gender gap of health care for the elderly can be extracted using the health related source data.
 - The narrower scope of definition can be set to ‘production of big data for women and families through various combinations of data generated by departments and agencies related to the policy on women and families’.
- We have determined three categories as potential areas to be developed as big data for women and families
 - For first category, presumes the collection of data generated

during the women and family policy and research process.

- This refers to various combinations of data produced by the Ministry of Gender Equality and Family, related departments of local governments, and the Korea Women’s Policy Institute.
- The second category is data generated by including gender awareness perspectives in the production phase of big data or by attempting gender separation.
- By analyzing gender differences of the existing big data materials, new women and family big data can be generated

<Table 5> Expected Topics for Gender Analysis of Big Data Materials

Source data	Big data	Topics on fender–family issues
Transaction data	Credit/debit card and financial transactions data, distributor data	→ Ex) Gender difference in credit/debit card usage and amounts used
Tracking device data	Mobile location data for telecoms, App-based Commercial Vehicle GPS Data	→ Ex) Gender difference in geographical shift patterns
Sensor data	Satellite imagery, roads/weather/power sensor	→ Ex) Gender difference in occurrence rates by time and location captured by CCTV
Online behavior data	Online pricing information Online search, registration information	→ Ex) Gender difference in online search and registration usage
Opinion data	Open SNS data such as Twitter	→ Ex) Gender differences in language usage of SNS postings by subject
Administrative data	Public administrative data (administration, tax, welfare, health, housing, etc.)	→ Ex) Connection between gender differences in tax and welfare

Source: Reconstructed based on Ji Young Lee (2015) p.53; recited on You Kyung Moon et al. (2016) p.18

2. Women's Policy Relations in Big Data Policy

- The Moon Jae-in administration, launched in May 2017, actively makes use of big data to reflect the voices of people more effectively.
 - It also presents specific policies based on big data analytics.
 - It take steps to utilize big data in the funding plans for policy implementation. For example, through the expansion of big data analysis function based on the National Tax Service (NTIS), the government makes efforts to minimize tax evasion income, by providing various and accurate reporting materials to enable taxpayers to file accurate reports.
- Each department has begun to actively utilize big data and establish long-term use plans for the use of such.
 - Finding the demand for policies in line with the characteristics of each ministry through big data, and is gradually increasing its use in the process of establishing and implementing policies.
 - Establishing an intelligent government based on recent technology such as artificial intelligence and big data, and to develop innovative financial services such as big data FinTech, it is utilizing big data to implement policies and provides human resources and material support accordingly.
 - Utilizing big data to solve problems in everyday life, such as early detection of children at risk and protection of the public from various disasters and safety accidents, and visible results are also being shown.
- From the perspective of gender awareness, the government's big data policy lacks specific policies on what kind of data is needed

to establish proper women and family policies and on how to generate the required data.

- The Ministry of Gender Equality and Family has no plans related to big data, and relevant policies of the National Statistical Office, the Ministry of Government Administration and Home Affairs, and the Ministry of Information and Communication are currently lacking policies on women or family related perspectives.
- Policies to foster human resources for active participation by female scientists and engineers in the big data industry are necessary.
- Comprehensive and long-term measures to utilize big data in terms of women and family perspective and to promote the female workforce are needed.

3. International Organizations and Overseas Cases

- In the 4th International Open Data Conference in Spain in 2016, the issue of female representation in the open data community and the need to improve gender balance in the open data center in 2018 is discussed.
- The U.S. Federal Trade Commission pointed out that big data analytics can impact and disadvantage female job seekers. To prevent potential discriminative actions, they suggest the followings; 1) review of proper representativeness 2) review of the biases in data processing algorithm such as big data collection and analysis 3) review of potential advantage and disadvantage of decisions made thru big data analysis and the negative effects on the specific demographic groups, and 4) review of fairness and ethical issues in

the utilization of big data review, including whether the potential advantage of the previously under-represented population is promoted.

- The State Personal Information Protection Committee in Ontario, Canada, announced in May 2017 guidelines on the use of big data such as quality, biases of data sets, and to prevent discrimination against factors such as race, nationality, color, religion, mental and physical handicaps in the process of big data analysis.
- In countries like Nigeria, high resolution GPS data with gender difference indexes is generated from the existing data on the growth rates of girls, promotion of women's literacy and contraception. Also, women's expense and movement pattern analysis using credit cards and mobile phone data for major Latin American cities, and mental disease identification thru the SNS analysis of the women and girls in India, South Africa, Great Britain, and US are the examples of using big data for promoting women's health.
- The Women's Big Data Human Resource Forum in the United States is a community of women experts on big data and its analytical technology Launched in 2015 with the support of U.S. big data industries, it is engaged in a variety of activities with the goal of encouraging, connecting, growing and achievement of female experts in the data science field, with the goal of expanding female representation in the big data field by more than 25% by 2020.

III. Availability of Public data on Women and Family Policy

1. Background and Purpose of Research

- Need to make efforts to maximize public big data, including administrative data, in line with the changing IT and social environment of the big data era
 - Big data has emerged due to factors such as the explosion of unstructured data due to the development of social network services, the decline of data storage and processing costs through cloud computing, the universalization of large - and high-speed networks, the increase of shadow information (GPS data, search patters, access data), and various background factors such as expansion of the information network of the object.
 - Public data held and managed by each institution will contain a wider variety of information through connection with other data, and its value will also increase, and is now being used in the public sector.
- Through this research, we will analyze the status of big data production and utilization regarding women and family issues such as gender equality, workforce development, human rights protection, and attempts to find new public data to support effective development of women and family policies.
- The scope of women and family big data in this study is based on the narrower scope generated by various combinations of data generated by relevant departments and agencies, and is expanded to

the broader scope that includes all big data analyzed with the gender awareness point of view.

- The broad scope is defined as all big data with gender awareness perspective, that is reanalyzed statistics relating to women and families in the existing big data.
- The narrow scope is defined as ‘big data produced through various combinations of data generated by departments and agencies related to the policy on women and families.’

2. Status of Public Data for Women and Family

- There are 184 survey statistics produced by research and development budgets in the women and family fields with only 14 (7.6%) nationally approved statistics.
 - The Ministry of Gender Equality and Family has 596 supported cases of research and registered in the Government Policy Research Information Service & Management (PRISM) from 2001 to the present (July 2017), and 574 cases are open to public.
 - 152 total cases, excluding research on children and adolescents, are categorized into gender equality (63 cases), family (37 cases), workforce development (28 cases), human rights protection (20 cases), and miscellaneous areas.
 - 98 reports (quantitative and qualitative) have been produced with more than one survey statistics, and 184 sets of survey data have been produced.
- The Ministry of Gender Equality and Family has seven information systems for the public and six for the business.

- Systems for the public include ‘ministry homepage’, ‘the e-history of Japanese military sexual slavery victim’, ‘History of Women Exhibition Hall’, ‘Sex offender alerts e’ and ‘child care services’ ‘female human resources use and gender equality implementation TF’, ‘The Ministry of Gender Equality and Family digital library’.
 - Systems for businesses include ‘Proposal management system for the Ministry of Gender Equality and Family’, ‘the gender impact evaluation system’, ‘the e-new job system’, ‘the female human resources database’, ‘the regional solidarity for children and women’s safety’, and ‘the preventive education management system’.
- The administrative statistics of the Ministry of Gender Equality and Family are based on one set of national approved statistics and 44 non-approved statistics.
- Non-approved statistics are classified into 7 cases of gender equality, 3 cases of workforce development, 20 cases of human rights protection, and 11 cases of family area.

3. Analysis of the Availability of Big Data for Women and Family

- Examination on how big data can be used by combining the government’s 100 new agendas, the 2017 Ministry of Gender Equality and Family Affairs’ agendas, and projects for the second half of 2017.
- Analysis of the use of Women’s Safety Big Data
- 19 cases of public data available, including information on sex offenders’ residential information.

-
- By utilizing this data, analysis of locations of violent crimes against women for improvement of women's safety, analysis of high crime areas with a large number of single female households, analysis of regional characteristics by type of crime, and analysis of public replies on certain issues posted on portal sites or SNS can be conducted.
 - These analyses can be used for policy development such as strengthened residential safety, assessment of the local government by the indicators of safety enhancement of vulnerable household, proactive response to violence against women, and research on the needs of policy consumers.
 - Available big data analysis for women with career interruption
 - 12 cases of public data available, including location information of new employment centers for women
 - Information input segmentation, systematization of e-new job center, and diversification of inflow channels to the new job center are needed to increase service utilization rate.
 - Potential use for preparing customized employment support measures for women with career interruption.
 - Available big data analysis for child care
 - 14 cases of public data available, including survey of gender equality and family status
 - Need to upgrade child care service and user systems
 - Advancement of child care service information systems, plans to enhance child care service user and provider matching, development of child care service types according to regional

characteristics, and preparation of follow-up management plans, collaboration of the Ministry of Gender Equality and Family Affairs with other related institutions, renovation of the child care service environment through IoT are available.

4. A Pilot Study of Big Data with Women's Family Public Data: Analysis of the Demand and Supply of Child Rearing and Care

- Using various public data utilization and collocation on care environment for the children aged 0 to 12 years old, the pilot cluster analysis is executed through the demand and supply indicators of child rearing and care environment of 230 different regions across the country.
 - Total 65 analysis indicators utilized as clustering variables
 - 5 demand indicators for child care services implemented by the Ministry of Gender Equality and Family (number of child care hourly/day service usage, etc.), 5 indicators of supply (child care service status, hourly/daily worker activity status per month, number of institutions providing service)
 - The demand indicators for the Ministry of Health and Welfare are 6 (number of infants and supply and demand for Family Care Benefit), 14 (Number of child care facilities per 100 infants, Number of national childcare facilities, and supply rate of child care facilities)
 - The demand indicators for the Ministry of Education is 7 (numbers of kindergarten and elementary school students, number of after school care class user), supply indicators is 14 (number of kindergarten, elementary schools, and number of schools with

afterschool care class, number of kindergarten elementary classes, number of afterschool care classes)

- Policy environment variables leverage 14 variables such as population, number of households, and regional gross product (GRDP)
- Key analysis result is as follows:
 - Based on this variety of big data-based cluster analysis results, mid - to long-term planning can be implemented to reflect demand and supply according to regional situations and characteristics
 - Cluster A (41 regions) Seoul and most cities in Gyeonggi Province
 - Cluster B (34 regions) Major provinces of Metropolitan cities and small and medium sized cities
 - Cluster C(47 regions) Old town of large cities with declining populations and small and medium-sized cities
 - Cluster D (47 cities) Urban or rural complex or rural areas
 - Cluster E (34 cities and provinces) Major population density areas
 - Cluster F (26 cities) Urban and complex area or rural areas with some demand/supply of child care
- Data is becoming increasingly important in today's society, and decisions based on data are becoming essential for businesses, policymakers and researchers. Big data has become a key concept in the use and analysis of data in societal development.
 - Women and family policy planning and research should also be advanced by generating research using various big data instead of relying only on the production and utilization of gender awareness statistics.

IV. A Study on the Public Opinion of a Childbirth Policy Using Data Science

- With solutions for birth rate issues being highly prioritized for local and national governments, evaluation of the experience of the direct consumers of childbirth and childcare services and related policies of the national and local governments is required.
- To that end, the study examines the difficulties facing major consumers of the delivery system through big data analysis and focus group interviews (FGI) on representative online communities of women in Korea, and attempts to suggest alternative measures.
 - For analysis of female life communities, a semantic network analysis is conducted by extracting keywords from postings about marriage, childbirth and pregnancy generated in 2016, from <lemon terrace>, <Momsholic>, and <82 Cook>.
 - To complement potential bias of online opinion analysis, and to understand the child care experience of full-time mothers, working mothers, fathers, and expectant parents, FGI for the directly involved in childbirth policies is executed.
 - Policy is proposed based on frames derived by integrating the online data analysis and the FGI interview analysis.

1. Online Big Data Analytics

- A total of 620,000 cases on birth and child care were collected and analyzed from <Momsholic>, <Lemon Terrace> and <82 Cook> in 2016.

- Some 540,000 of those cases were collected from <Momsholic>. <Lemon Terrace> takes up only 12% since its postings are generally about wedding process, and <82 Cook> is about one percent, since its most members are generally over the childbirth and care ages.
- As a result of a semantic network analysis on the entire postings, ① Career issues such as reinstatement, ② infant and toddler care institutions, ③ childcare, and ④ post-natal care are most prioritized issues, all of these issues are linked to ‘husbands’, and expectations and disappointment regarding the spouse are mentioned over and over again throughout the maternity process as well as mentions of feelings of regret and guilt about the child, family, husband, etc.
- On the ‘childcare’ topic, the starting point and the most frequent keyword appear to be ‘husband.’
 - The spouse’s contributions to childcare and housework seem to be low, and depression caused by repeated childcare and household chores is also frequently expressed.
 - Unhappiness about husbands’ inability to act as an intermediary in relationships with in-laws, as well as difficulty in emotional investment for relationship intermediation amidst a busy lifestyle
 - Parental leave from work is linked to the possibility of quitting, and many moms are rushing to return to work due to the perception that husband’s income alone is financially insufficient.
- The semantic network on ‘working mother’ reveals that there is no public entity in the surrounding environment to help moms with childcare between the company and the family.

- Results show that many people express stress about their careers, such as paternity leave and reinstatement, and career related discussions such as “company” and “commute” are converged towards difficulties in sharing the household duties with “husband”.
- The guilt of having to send a child to the ‘all-day’ class of daycare center, as well as various concerns such as the shortage of national and public daycare centers, treatment of the teacher, and abuse cases are expressed.

2. FGI Analysis of the People Affected by Policy

- Child care and childbirth policy consumers are divided into five categories: (1) full-time moms, (2) working moms, (3) expectant moms (4) DINKs and (5) single income dads, and total of 19 persons, with 4 for each group (except for the expectant moms which are 3 persons) were interviewed and analyzed with semantic network analysis technique.
- Full-time mom - inevitable career interruption
 - Career break of the mom is generally determined by ① whether “the mother’s mother ” or caring worker can be used in childcare, ② whether it is guaranteed to resume the job after maternity leave and ③ if the spouse respects mom’s professional activities and is free from fixed gender role custom.
 - Most housewives become “full-time mothers” after childbirth by ‘forced choice’ due to the above “hierarchical” conditions, and experience feelings of alienation from social relationships, and loss of self-identity.

- The social perception of them as “leisurely” housewives also adds to feeling of alienation, and the social infrastructure that is not considerate of the families of pregnant women and infants often makes matters even worse.
- Working moms - between career and guilt
 - The essential condition for work and family compatibility is the ability to afford a female caretaker for their children. Even with the help of their in-laws, their mothers, or the hired caretakers, mothers are still in charge of caring for them.
 - Results indicate that still only a small number of men use paternity leave, and the general social myth and social system that only mothers are considered as major caretakers must change.
 - Ultimately, with a chronic “lack of time,” working moms feel guilty and indebted, as if they are insufficient in all aspects of work, family, and childcare.
- Expectant mothers - confusion and worrying
 - They are most concerned about career breaks due to childbirth and childcare, and ‘worry’ rather than guilt or depression is expressed in uncertain circumstances after pregnancy
 - Support is provided in the form of a pregnancy voucher, the ‘Good Heart Card’, but most of the bills are spent up to the birth and within 30 weeks, and it causes suspicion and complaint that the hospital may undergo unnecessary examinations.
- Single-income Dads – Doing their best and being misunderstood
 - Although there is a sense of pressure about being solely responsible for family finance, it is perceived as an “econo-cultural destiny.”

- They consider the spouse's career break as inevitable or voluntary, which is in contrast to mothers' perceptions.

○ Non-marriage/DINKs – Aversion to risk

- A rejection of the 'risk' inherent in carrying out childbirth under socioeconomic uncertainty, including the health of the fetus, the birth of the baby, the family, the economy, and the return to the labor market after the birth.

3. An Analysis of Major Presidential Candidates' Campaign Pledges

- In the 19th presidential election, a total of 76 slogans and policies were presented in the official pledge lists of five major candidates regarding childbirth and child care, and a consensus on the low birthrate being problematic has been confirmed.
- Although semantic network analysis showed that major directions of policy meets the needs of the mothers to ease working conditions and strengthen public care, the specific aspects of each party are quite different.

4. Conclusion

- The class deviations that determine different child care and welfare conditions strongly affects the difficulties women experience in their pregnancy, childbirth, child care and reinstatement efforts, making the lower economic class more vulnerable.
 - Despite the fact that it these are social issue, childbirth and childcare are considered "personal" issues and parents are required to bear all the responsibilities.

-
- A partial support policy such as part-time free childcare and pregnancy voucher alone cannot control the “markstreaming” of childbirth and childcare, and thus continued perception that child birth and care is high cost makes the issue of low birth rate hard to address.
 - Paradoxically, the difficulties that converge into needs of a husband’s role in the child care process show the absence of a public actor for childcare. Meanwhile, men face difficulty in living up to family expectations resulting from traditional Korean socioeconomic culture.
 - The social stereotype that childcare and housework belong to women, the labor environment in which maternity leave and parental leave are not guaranteed, and the practice that only mothers are required in the educational institutions need widespread overhaul.
 - Therefore, a more comprehensive maternity care policy should be formulated and enforced, involving changes in class and working conditions.
 - Korea’s male-centric work culture makes distribution of childcare difficult.
 - Beyond interpreting the difficulties of the mother’s birth experience as a gender issue, comprehensive policies are requested to transform the Korean social work culture and compensate for class deviations.

V. Analysis of the Survival Rates of Self-Employed Businesswomen Using Credit Card Big Data

1. Purpose of Research

- The current status of start-up/closure and survival rates of self-employed workers are analyzed by gender comparisons.
 - Self-employed workers have operational independence and autonomy compared to wage workers. On the other hand, they are less stable due to the disadvantages like long working hours, uncertain operations, and low profits.
 - Korea's self-employed businesses are generally small in size and low in performance efficiency, so countermeasures are required to ensure their stable operation. The female self-employed workers have less funds for starting businesses, less work experience and less specialized expertise than men. In addition, they demonstrate poor business performance and consistency due to the burden of caring for children under the age of six and the household duties.

2. Building Big Data Materials

- The data for analysis is based on information of individual business owners who are Shinhan Credit Card members from 2001 to 2016.
 - The variables for analysis are the gender of the business owner, the region (17 cities and 25 districts in Seoul), the business type (13 sections, 57 detail types), the date of starting business, the date of closure, and whether it is still in business as of year 2016.

- Unlike previous cross section surveys, this data identifies the status of closure through time series analysis and could include nearly all of the target population.

3. Analysis Results

- According to our analysis, the number of startups run by women is higher than that of men.
 - However, the number of business closures is higher too, so the difference in number of businesses staying in operation by gender gets lower than that of opening start-ups.
 - Over the past 15 years, the number of entrepreneurs is 3.14 million female and 2.35 million male, with about 790,000 more female start-ups.
 - As of December 2016, the number of businesses that remain active is 920,000 for women and 750,000 for men, decreasing the difference in number by gender to 170,000.
- Changes in the number of entrepreneurs by year have shown similar trends by gender and age.
 - The number gradually decreased from 429,000 in 2002 to 278,000 in 2004. Then it increased to 394,000 start-ups in 2009. Since then, numbers have remained steady.
- The five-year survival rate of self-employed businesses is relatively low at 29%.
 - In terms of gender, female-run businesses are 4.4% lower, with 27.1% for women and 31.5% for men.

- Difference in gender survival rate gradually decreased across recent years, from 7.0% in 2002 to 2.2% in 2011.
- Across 13 business sectors, both men and women have the highest number of startups in the restaurant industry; for women the next highest are apparel and beauty industries, for men the next are domestic/ living services and food/beverage businesses.
 - As for the survival rate for 57 specified business types, women's survival rates are the highest at 48.8% for general medical practices, followed by the home design business and jewelers. On the other hand, the ocean farming industry was the lowest with 12.9 percent, while the entertainment establishment industry at 14.1 percent, and the nightlife industry at 15 percent.
 - For men, home design businesses showed the highest survival rate at 57.9 percent, followed by dry-cleaners and auto repair service. The lowest survival rate is in the nightlife industry and other entertainment establishments at 9.6 percent each.
 - The industry with the highest rate of survival for women compared to men is beauty businesses (14.2% difference), followed by the nightlife and cosmetics industries, which are 5.4% and 5.3% higher. The characteristics of these industries are that women are more likely to start businesses than men, but generally have a lower survival rate. These patterns demonstrate need for improvement.
- With regards to age, women in their 40s are the most active in starting business, compared to men in their 30s.
 - Survival rate by age increases for both men and women in their 50s

- Survival rates by age vary over time. When comparing data from 2002 and 2011, women showed an increase in survival rates across all ages, with greater increases amongst younger age groups. Meanwhile men's survival rates increased for those in 20s and 30s between 2002 and 2011, while 40s remained consistent and 50s and 60s decreased. In short, the gap in survival rates by age is decreasing.
- With regards to region, new businesses are most established in Gyeonggi and Seoul.
 - For survival rate by region, Sejong is the highest at 34.5% and Gwangju lowest at 25.4%.
 - Survival rates are higher for men than for women in all regions.
 - Gangnam district has the largest number of women business owners with 52,000 and men with 48,000, four times the size of Seongdong-gu, which has the smallest number.
 - Jongno district has the highest survival rate of the female self-employed at 33.4 percent, and Gangnam district is the lowest at 23.0 percent.
 - Gangnam district has the highest start-up rate and the lowest survival rate; the difference by gender is also the smallest in Gangnam.
- New research methodologies and efficient data integration of big data remain a challenge.
 - It is difficult to obtain all the necessary variables for big data analysis by collecting and reprocessing raw materials generated for other purposes. New research paradigm is needed.

- An example of efficient data combination: variables affecting the female self-employed can be discovered by linking household/population surveys with self-employed business data such as family status, working conditions, and managerial situations; this leads to a new dimension of analysis.

VI. An Analysis on the Status of Re-employment of Women with Career Interruption Using Public Big Data

1. Purpose of Research

- For a long time the government has been making various attempts to improve the participation rate of women in the economy. Since 2009, the government has set up about 150 new job centers for women (Saeil Center) nationwide to help women with interrupted careers, and provide various types of support for women seeking to rejoin the workforce.
- While the number of companies and job seekers who use the Saeil Centers is increasing exponentially, the characteristics of the companies that are seeking employees through the Saeil Centers and those of women who are trying to find the job have yet to be fully identified. Furthermore, there is still a lack of assessment of the jobs that are posted on the Saeil Center. Further close analysis is required to ensure the effective operation of the Saeil Center.

2. Building Big Data Materials

- Recruiting and job search information for the Saeil Center can be uploaded and searched through the centers' online system. The system was built mainly for efficient operations of the center, but information accumulated therein can be converted for big data analysis.
- This study has extracted data about participants, such as location of companies, type of industries, positions, numbers of personnel, and required education, as well as data about the jobseekers, such as the age, education, location, type of jobs wanted, and positions applied to in order to build big data source for analysis from 2014 to 2016.

3. Analysis Results

- First, a growing number of companies are using the Saeil Centers for recruiting. In 2014, about 120,000 companies recruited more than 300,000 employees, and in 2016, the number of recruiting firms reached 190,000 and the number of hires increased to 420,000. Although these numbers have somewhat plateaued recently, companies that post job ads at the Saeil Centers were found to recruit about 2.5 people per posting. Also, the number of job seekers using the Saeil Centers increased by 37.2 percent from 280,000 in 2015 to 390,000 in 2016. As the number of Saeil Centers is increasing, the number of businesses and job seekers using them is increasing as well.
- Second, it appears that the industries that have hired workers through the Sail Center were most concentrated on the traditionally women's jobs of "health and social welfare services" In 2014, these two industries (by mid-division) accounted for 29.1% of all recruiting

firms, and went up to 31.2% in 2015 and to 31.6% in 2016. Meanwhile, the share of “manufacturing industry” is gradually decreasing from 21.4% in 2014 to 12.4% in 2015 and 20.1% in 2016.

- Third, it is found that the overwhelming majority of firms recruiting through Saeil Center are small businesses with less than 50 employees across all regions in the nation. Considering the gap in wage and welfare benefits between large companies and small businesses, this suggests that quality and level of jobs recruited through Saeil Center may be relatively low.
- Fourth, according to the classification of Korean standard occupations, most jobs posted in the Saeil Center are found to be “office workers,” “service workers” and “manual laborers.” Around a quarter of all jobs are office workers, another quarter to be the service workers and manual laborers occupy a sixth. However, the ratio of office workers (24.4 percent) was overtaken by that of service workers (27.4 percent) in 2016, resulting in a reversal of the ratio between the two jobs. Meanwhile, the ratio of manual laborers is increasing. Therefore, the industrial compositions and occupations offered by recruiting companies seem to be declining.
- Fifth, the percentage of companies seeking permanent jobs is decreasing sharply, although there is a significant variation in the number of attempts across the regions. The percentage of companies seeking permanent employment is 65.5% in 2014, 64.1% in 2015 and 62.3% in 2016. As of 2016, 41.4 percent of companies in Sejong City were found to have the lowest percentage of full-time workers, 48.7 percent in Incheon, and 75.8 percent in Gwangju.
- Sixth, the most desired jobs are managerial/accounting/office jobs, with 24.7%, or about a quarter of all job seekers hoping. Health

and medical related jobs were following at 12.4 percent. Third, many women want jobs related to social welfare and religion.

- Seventh, the higher the training level of job seekers, the more likely they are to work outside of their area of residence. This means that the higher education levels implies more investment of resources in human resources and therefore higher opportunity cost, resulting in higher willingness to work in other areas.
- Finally, the average age of foreign women using the Saeil Center was lower than that of Koreans. The majority of Koreans were concentrated in their early 40s to early 50s, while foreigners were found to be in their late 20s to late 30s. However, the level of education for foreign job seekers was found to be lower than that of Koreans. While high school graduates represented the highest percentage of both foreigners and Koreans, the percentage of high school dropouts among foreigners was twice as high as that of Koreans, while the percentage of college graduates was significantly lower.

VII. Policy Proposal

1. Policy Suggestions by Subject of Big Data Analysis

A. Public Opinion Analysis on the Birth Policies Using Data Science

- Reconstituting the social support and value of caretaker work
 - The most urgent policy and service development regarding the stratification of childbirth and childcare is social support of caretaker work.

- In addition to the expansion of public daycare centers and kindergartens, providing support to those who hire professional caretakers at home should be considered.
- Re-establishment of the value of caretaker work is also necessary for effective policy/service implementation.
- Using the system of maternity leave, paternity leave, shorter working hours, etc.
 - In large corporations, public institutions and professional jobs, women's maternity leave, parental leave, and childcare facilities in the workplace are relatively well utilized, and employees are less likely to experience practical or relational discrimination when they return to work after childbirth, resulting in balance life between family and work.
 - When necessary, punitive measures against companies should be concurrently available.
- Raising communal awareness about mothers and children
 - Providing services for the expansion of small community networks for full-time mothers, within a social environment is isolated or within a social environment unsafe or unacceptable to children, is expected to help reduce the personalization problem of childbirth and childcare.
 - An effective example would be small community libraries operating at 6,334 locations nationwide.
 - Facilities need to be expanded to be more convenient for children and mothers to use public transportation and public facilities.

-
- Modification of the voucher system
 - Obstetricians tend to intensify anxiety in mothers who worry about their unborn child's health and abnormalities and often demand excessive examination and treatment.
 - Most 'Good Heart Cards' are depleted before the birth, and a single blood test required for older pregnant women would use up the value of the cards.
 - Mandatory paternity leave
 - Implementation would not only relieve the burden of working mothers and full-time mothers, but would also play a key role in easing outdated gender practices throughout society.
 - Education about childbirth and childcare for fathers needs to be promoted as well.
 - Adjustment of 'normal family' paradigm
 - There needs to be a paradigm shift in which low birth rates are not just considered in regards to 'normal family' units but to include single mothers and DINKs.
 - In order for single mothers' work and family life to be compatible, it is necessary to socialize a significant portion of their childcare and to alleviate the gender structure of the labor market in which two-member families can live comfortably with the wages of women alone.

B. Analysis on the survival rate of the self-employed women by credit card big data

- Need policies for startup and sustainable operations
 - The five-year survival rate only stands at 28 percent, indicating that the self-employed are unstable in long-term operation.
- Redirection into industries with high survival rates
 - Businesses with high survival rates tend to require professional expertise or large capital investment. Meanwhile, those with lower barriers for entry such as the farming industry, leisure industry and bar industry tend to have low survival rates.
 - Sectors with higher survival rates for women than for men have lower overall survival rates when both men and women are counted than for other industries. In order to increase the survival rate for women, businesses that have high survival rates should be introduced through help with acquisition of expertise and the preparation of capital prior to their launch.
- Need for policies to take into account characteristics by gender and age
 - The business survival rate of males in their 50s and 60s is decreasing, which are the major age groups of the self-employed.
 - The effect of new technology on the development of science and technology is more advantageous to the younger age groups and yet is presumed to become more important in business survival, while differences in career status or accumulation of pre-business experiences are less likely to affect survival.

- Additional analysis of these characteristics by gender and age and countermeasures are accordingly required.
- Establishment of big data linking personal information with franchise merchant information
 - If variables known to affect performance of self-employed women (such as presence of a child under six years of age, the level of burden of household care, household variables such as income, demographic variables like marital status and education, and employment variables such as work experience status) are linked to performance variables of franchise merchants, more intensive analysis could be conducted.
- C. Study on the Status of Recruiting Career Interrupted Women Using Public Big Data
 - Securing adequate jobs
 - The amount of jobs posted at the Saeil Center is increasing over the years, but the quality of the jobs is getting worse. In order for the Saeil Center to provide substantial help to those with career interruption, more adequate jobs need to be provided.
 - Consulting of jobs targeted to job-seeker's characteristics
 - Consulting for jobs based on the characteristics of the job-seeker is important, since the type or quality of the job desired may vary depending on the socio-economic conditions of the job-seeker (in particular, level of education or nationality).
 - A study on how to reflect the needs of job seekers is needed as a wide disparity between the jobs companies needs to fill and the job prospective employees want has become apparent.

- Securing data for analysis of re-employment of job seekers
 - While analysis on successfully re-employed applicants and the amount of time invested for reemployment is necessary, such research cannot be conducted due to personal information protection issues.
 - If data can be obtained in the future, we expect to find out how the chances of reemployment vary depending on individual characteristics.

2. Policy Proposals for Big Data and Production

A. Plan for utilizing big data on Women and Family

- Review of how big data can be used by combining the 100 new government agendas, the 2017 agendas of Ministry of Gender Equality and Family Affairs, and their major projects in second half of 2017.
- Ideas to utilize big data on female safety
 - 19 cases of public data available, including information on sex offenders' residential areas
- Ideas to utilize big data on career-interrupted women
 - 12 cases of public data available, including location information of Saeil Center for women
- Ideas to utilize big data on child care service
 - 14 cases of public data available, including survey of gender equality and family status

B. Plan on Big Data Production with Gender Awareness Perspective

○ Collection and production of gender-conscious data

- Gender-conscious data is a concept that extends the scope of gender statistics and refers to data that can be used for gender equality or to show gender characteristics.
- The UN Statistical Office's Data2X Initiative is developed to clarify and fill the gaps in gender-conscious data.
- In Korea, government organizations or programs for collecting and producing gender-conscious big data have not yet been implemented. More future efforts are required in this area.

○ Expansion of women's presence in big data fields

- Growth in the big data industry provides new employment opportunities for women.
- The 4th International Open Data Conference discussed female representation in the open data community.
- The Women's Big Data Human Resource Forum in the United States is a community of women experts in big data and analytical technology, with the goal of expanding women's representation by more than 25 percent.

○ Strengthening big data research with a gender-conscious perspective

- Research is needed to find issues where big data analysis can work effectively for women and family policy studies and to review the production methods of big data to promote such research.
- Research to replace and supplement existing statistics should be conducted; particularly, surveys with sensitive issues such as violence against women, prostitution and multiracial families.

-
- Evaluation of the gender impact of public and private sector big data business
 - Need to evaluate the entire process of public and private big data projects, from the planning phase to produced results, from a gender conscious point of view
 - Big data guidelines published by the Personal Information Protection Committee in Ontario, Canada, require vigilance against discrimination during the collection, integration, and analysis phases.
 - Identifying and improving the status of gender discrimination in the utilization of big data such as artificial intelligence
 - While there are positive aspects of big data analysis, such as the development of new products and services, these processes may be detrimental to low-income and impoverished households.
 - Enterprises in the United States have an obligation to ensure that their utilization of big data analysis complies with anti-discrimination laws.
 - Need to identify and improve the status of gender discrimination in the current use of big data in order to prevent possible discrimination.



KWDI

Korean Women's Development Institute

225 Jinheung-ro, Eun pyeong-gu
(1-363, Bulgwang-dong) Seoul, 03367, Korea
TEL 02.3156.7000 FAX 02.3156.7007
<http://www.kwdi.re.kr>