

2019

Research Paper-28



# Improvement and Discovery of Policy through Utilization of Gender Big Data (II)

You Kyung Moon  
Ki Taek Jeon  
Young Taek Kim  
Sungmi Jung  
Ho Joong Bae  
Hee-Tae Chung  
Youjing Kim



Korean Women's Development Institute



## Improvement and Discovery of Policy through Utilization of Gender Big Data (II)

---

©2019

Korean Women's Development Institute

225 Jinheung-ro Eunpyeong-gu

Seoul 03367

Korea

[www.kwdi.re.kr](http://www.kwdi.re.kr)

# Contents

<b>I. Introduction</b>	<b>1</b>
1. Purpose and Content of Research	1
2. Research Methodology	3
<b>II. Background of Big Data Research on Women and Family</b>	<b>5</b>
1. Concepts and Categories of Big Data for Women and Family Sectors	5
2. Relationship of Big Data Policies and Women and Family	6
3. Case Studies on the Structure and Utilization of International Organizations in Major Developed Countries	7
<b>III. Potential Gender Discrimination in the Utilization of Big Data and Possible Countermeasures</b>	<b>7</b>
1. Relevant Issues	7
2. Debating the Neutrality of Big Data Algorithms	7
3. Potential for Discrimination by Algorithms	8
4. Cases of Gender Discrimination by Algorithms	11
5. Measures to Prevent Discrimination in Big Data	12

<b>IV. A Study on the Utilization of Spatial Big Data in the Women's Safety Sector</b>	<b>12</b>
1. Background and Purpose of Research	12
2. Policy Trends and Examples of Usage for Spatial Big Data	13
3. Results of a Pilot Study on Spatial Big Data for Women's Safety	14
4. Methods of Utilizing Spatial Big Data for the Women's Safety Sector	15
 <b>V. Big Data-based Analysis of Female High-risk Drinking and Policy Alternatives</b>	 <b>15</b>
1. Background and Purpose of Research	15
2. Study on the Current State and Trends of Female High-risk Drinking	16
3. Study on the Current State and Trends of Female Liquor Consumption Using Credit Card Big Data	17
4. Study on the Effects of Female High-risk Drinking Using Health Insurance Industry Big Data	18
5. Suggested Policies on Female High-risk Drinking	19

<b>VI. The Apparent Impact of the “MeToo” Discourse in Social Big Data</b>	<b>19</b>
1. Necessity and Purpose of Research	19
2. Key Research Results	20
3. Implications of Policies related to the “MeToo” Movement	23
 <b>VII. Proposed Policies for the Production and Utilization of Big Data</b>	 <b>24</b>
1. Establishing Production Systems of Big Data for Women and the Family	24
2. Measures to Prevent Discrimination in Big Data	26



## Tables

〈Table 1〉	Commissioned projects and involved institutions .....	4
〈Table 2〉	Seminars on Women and Big Data: Gender Issues in Big Data Research Methodology .....	4
〈Table 3〉	Presented topics at the “Women and Family Big Data Symposium” .....	5

## Figures

[Figure 1]	Causative elements of potential discrimination in an algorithm .....	9
[Figure 2]	Content Analysis of the Concepts in Postings .....	21
[Figure 3]	Content Analysis of Comments, Shares, and Replies ....	22
[Figure 4]	Proportion of Different Influencer Groups' Comment and Re-comments by Month .....	23



# Improvement and Discovery of Policy through Utilization of Gender Big Data (II)<sup>1)</sup>

You Kyung Moon

Ki Taek Jeon

Young Taek Kim

Sungmi Jung

Ho Joong Bae

Hee-Tae Chung

Youjing Kim

## I . Introduction

### 1. Purpose and Content of Research

- The widespread development of the computer technology has given rise to the generation and analysis of big data, combining information from disparate and diverse sources in ways that were previously impossible. Accordingly, both the industrial value of big data and the necessity of relevant policies are growing by the day. As such, we can expect that research on policies regarding women and family sectors will also expand to encompass new topics and results that were unavailable through existing research methods.

---

1) This study is a collected report as part of collaborative research series of National Research Council for Economics, Humanities and Social Studies, and includes content of two commissioned projects. Refers to Appendix1 for details.

- Big data research regarding women and families is expected to consist of two general categories. First, is the foundational research of big data itself. It is necessary to study the overall potential for utilization and availability of big data regarding women and family policies, including what sort of data is needed for relevant research, which methods and analysis processes are necessary for the production of such, as well as potential future applications. Second, in addition to foundational research, understanding and addressing the issues present in the women and family sectors will require a pilot study in advance of implementing a full-scale study. Big data analysis has great potential for efficacy and by aggregating and processing big data materials and pursuing effective analytical methods in service of an immediately relevant research topic, we expect this study to yield significant implications for the future of big data research on women and the family.
- This study is a long-term project spanning 3 years (2017-2019). The primary overall goal of the project is to develop methods of improving, managing, and sharing data in order to more effectively facilitate the use of big data in research and policy decisions regarding women and the family. Secondly, this study intends to examine possible utilizations of big data by conducting a trial study on certain policies that demonstrate widespread disparities in beneficial effectiveness or significant blindspots, yet nevertheless have significant potential applications for big data. Finally, the results of this trial study will be used to derive methods of improvement for women and family policies. In 2018, the second year of this study, we first reviewed policies for developing female-centric big data in international organizations and developed

countries in order to establish a basis for our first primary objective of applying big data to women and family policies. Secondly, we explored how spatial data, a highly applicable type of big data, can be utilized in studies regarding women, and thirdly, we examined the possibilities for gender discrimination in the utilization of big data and countermeasures thereof. With regards to our second primary objective of a selective pilot study, we first attempted an in-depth analysis of the current state of high-risk drinking behaviors by women and countermeasures against such through incorporating pre-existing big data such as population statistics, credit card data, and public health information. Secondly, we attempted to examine the efficacy of using spatial big data comprising of information on spaces, facilities, and locations, to illustrate the safety statistics of women. Finally, we interpreted the widespread implications and impact that last year's "Me Too" movement had upon the discourse of social data.

## 2. Research Methodology

- Aggregate domestic and international literature and data
- Aggregate public and private big data usages cases related to women and families
- Review big data raw materials by category and attempt to identify and analyze potentially relevant materials
- Commission necessary tasks to experts in relevant fields and further communicate with big data specialists in the form of industry-academic collaborative research.

〈Table 1〉 Commissioned projects and involved institutions

Project 1: Big Data based Analysis of Women's High-risk Drinking and Policy Alternatives (by Inje University)
Project 2: Plans for Utilizing Spacial Big Data in Safety for Women (by Korea Research Institute for Human Settlements)

- Social network data was analyzed in collaboration with private big data company *Ars Praxia*.
- Four interdisciplinary seminars were held by multiple institutions to enhance female researchers' understanding of big data and to promote cross-disciplinary informational exchanges.

〈Table 2〉 Seminars on Women and Big Data: Gender Issues in Big Data Research Methodology

	Subject
1st	Business Innovation thru Nonlinear Big Data Aalysis
2nd	Traffic Big Data Status and Road Congestion Forecast System, Transportation and Women's Safety
3rd	Building e-Childen's Wellness Support System using Big Data
4th	Measuring Living Population of Seoul and Policy Development

- The “Women and Family Big Data Symposium” was held to share the results of a year of research with female researchers and policy makers, and to provide a platform for discussions among experts.

〈Table 3〉 Presented topics at the “Women and Family Big Data Symposium”

	Subject
Presentation 1	Analysis of Spreading “MeToo” Discourse usng Social Big Data
Presentation 2	Big Data-based Analysis of Womens’s High-risk Drinking and Policy Alternatives
Presentation 3	Plans for Utilizing Spacial Big Data for Safety of Women Sector
Presentation 4	Potential Gender Discrimination in the Utilization of Big Data and Countermeasures

- Attended major big data seminars and forums in Korea, information sessions, and advisory meetings.

## II . Background of Big Data Research on Women and Family

### 1. Concepts and Categories of Big Data for Women and Family Sectors

- Academic and social discourse on the concept of big data in women and family policies is currently in its early stages. This study conceptualizes big data on women and the family as the categories of big data necessary for research and policy changes on topics such as gender equality, gender disparities, improvement of social conditions for women, and gender equality in the family.
- The concept of ‘big data on women and the family’ can be further organized into two major definitions. On a macro scale, it can be understood as “the reproduction and re-analysis of all big data from a gender-conscious perspective.” In other words, this entails a

gender-conscious re-construction of processes from collecting data to interpreting results with attention to issues such as gender discrimination, gender differentiation, and female-specific conditions. For example, if big data from the National Health Insurance Service is used to further generate big data on gender differences in elderly populations, the latter generated data can be considered as big data on women and family. On a narrower scale, the concept can be defined as “the generation of big data on women and family through various collaborations of resources produced by public institutions and organizations related to women and family policies.”(Moon et al. 2017, p.16)<sup>2)</sup>

- Production of big data on women and family under this specific definition requires active interest and efforts of relevant ministries and organizations such as the Ministry of Gender Equality and Family and the Korean Women’s Development Institute are necessary. In other words, certain process of operation can be set up so that the data generated in the processes of policy and research on women and family can be designated as source data of this type of big data.

## 2. Relationship of Big Data Policies and Women and Family

### A. Big data policies for government projects

- Neither 「The Moon Jae-in Administration’s 5-year Plan for Government Operation」 (Advisory Committee on State Affairs Planning, 2017) nor the “Public Data Innovation Strategy” published in 2018 include big data policies specific to women and family.

2) Re-written based on You Kyoung Moon et al. 2017. *Policy Innovation and Development Using Big Data on Women and Family*, pp.15-18, and You Kyoung Moon et al. 2017. *Basic Research for Utilization of Big Data on Women and Family*, pp.7-18.

### 3. Case Studies on the Structure and Utilization of International Organizations in Major Developed Countries

#### A. Current utilization of big data by the U.N.'s Data2X<sup>3)</sup>

- Data2X is a technology and support platform developed for the quality and utilization of gender-related data with the aim of improving the quality of life for women around the world
- Data2X utilizes big data typically characterized by its high volume, diversity and quickness, to encourage research and studies focused on improving women's quality of life.<sup>4)</sup>

## III. Potential Gender Discrimination in the Utilization of Big Data and Possible Countermeasures

### 1. Relevant Issues

- The gradual universalization of big data algorithms results in some negative repercussions. As such, issues regarding the equity, neutrality, and transparency of such algorithms have emerged.

### 2. Debating the Neutrality of Big Data Algorithms

#### A. Debating algorithm neutrality in The EU

- Through the GDPR, the EU has stipulated its right to question the neutrality of algorithms that are applied to individuals and the resultant right to refuse the usage of some algorithms.

3) Gender Data Gaps-data2x. <https://www.data2xorg.what-is-gender-data/gender-gaps>. Connected on 2018.4.28.

4) Gender Data Gaps-data2x. <https://www.data2xorg.what-is-gender-data/gender-gaps>. Connected on 2018.4.29.

#### B. The World Wide App Foundation's Algorithm Accountability

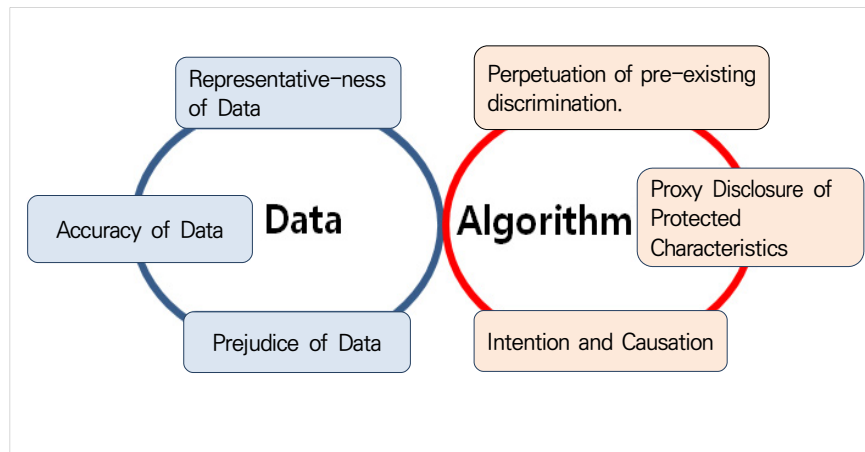
- 3W Foundation indicates that various algorithms can demonstrate bias and discrimination and suggests propositions regarding the causation and prevention of such.
- The Foundation identifies possible causes as biased or substandard input data, flawed application of rules, lack of contextual awareness, and repeated errors in operation.
- Suggested preventions include the principles of equity, understandability, verifiability, accountably and accuracy.

#### C. Debating Algorithm Neutrality in the U.S.

- The U.S. Federal Fair Trade Commission and the White House stated that big data algorithms present both opportunities and concerns, and that appropriate efforts must be taken to prevent those concerns from being realized.
- Big data algorithms have ambivalent tendencies, i.e. they are useful and valuable devices that are also potentially harmful to users, especially low-income and vulnerable populations

### 3. Potential for Discrimination by Algorithms

- The causative elements of a discriminatory algorithm can be generally distinguished as the algorithm's input data and the algorithm itself.



Source: Written by the author

[Figure 1] Causative elements of potential discrimination in an algorithm

#### A. Representativeness of data

- Big data analytics are often based on data sets that lack data from certain demographics or populations.
- Collected big data may not accurately represent the populations necessary for the purpose of analysis.
- Case study: Boston's Street Bump
  - Intended to contribute to improving road conditions by providing information based on smartphone GPS data
  - Lacked road information for low-income areas and focused on improving road conditions in high-income locales, thus inadvertently committing economic discrimination

#### B. Accuracy of data

- Big data generated by digital systems is often assumed to be accurate,

but in reality much of it is based on data with inaccuracies.

- Big data contains a lot of garbage data, thus it is important to refine its products.

#### C. Prejudice of data

- Since big data is data generated for a specific purpose, it is a dataset with pre-existing context and characteristics.
- When using big data, it is vital to understand these prior contexts and potential biases.

#### D. Perpetuation of pre-existing discrimination.

- Algorithms that have learned patterns from previous data can reproduce or reinforce existing patterns of discrimination.
- Case study: Google's targeted ad algorithm for criminal record information websites
  - Publishing ratio of ad links (Black Name > White Name) (Sweeney, 2013)
  - An algorithm with the commercial purpose of targeted advertising thus learns and perpetuates socially pervasive racial prejudices.

#### E. Potential for proxy disclosure of protected characteristics

- The United States legally prohibits discrimination based on protected characteristics (race, skin color, gender, religion, age, disability, country of origin, marital status, and genetics etc).
- Big data runs the risk of reproducing discrimination by generating proxy information for such protected characteristics. Furthermore,

developers of such algorithms can reasonably feign inadvertence (Ohm, 2014).

#### F. Intention and causation

- Big data analytics claim the ability to process vast amounts of data with such scale and efficacy that they detect patterns and relations that were not otherwise hypothesized (e.g. the relationship between beer and diaper purchases).
- Big data algorithms tend to assume that correlations are causative relationships, while causative assumptions can provide the basis for discrimination.

### 4. Cases of Gender Discrimination by Algorithms

- The patterns of gender discrimination in hiring practices and promotions have been reproduced by big data algorithms.

#### A. Case study: Google's job advertising algorithm

- A research team from Carnegie Mellon University (Datta et al.) developed a tool called AdFisher to prove the professional sexism of the algorithm (Datta et al., 2015:152).

#### B. Case study: Income estimation algorithms of domestic credit information companies

- Domestic insurance screening processes run the risk of unjustly rejecting female applicants since adopting and actively utilizing big data algorithms in September 2017.
- Female loss rates tend to be high in life insurance products, due

to the fact that female clients' loss rates are reported as high if their estimated personal income is below a certain amount (Jeong Jae-wook et al, 2010:11-15).

- These research results and the use of data-learning insurance screening algorithms indicate that such algorithmic processes can automatically result in discrimination, despite the regulations of insurance products technically prohibiting biased selling based on gender.

## 5. Measures to Prevent Discrimination in Big Data

- In this light, it is clear that Korea's public research institutes need to be at the forefront of examining extant big data algorithms for societal prejudices and devising specific guidelines for the life cycles of big data. Furthermore, the need for developing regulatory compliance services such as algorithm audits and external testing should be considered as part of the quantitative and qualitative growth of the big data industry.

# IV. A Study on the Utilization of Spatial Big Data in the Women's Safety Sector

## 1. Background and Purpose of Research

- Women's safety has garnered much attention recently due to increased sexual violence and high-profile misogynistic murders.
- The purpose of this study is to utilizing recently designated big data regarding transient populations and credit card information to

identify geographical and spatial factors and characteristics relevant to women's safety.

- Conclusions were drawn by examining the definitions, importance, and utilization trends of spatial big data as well as cases wherein spatial big data was effectively used for preventing crime or making arrests after the fact.
- A pilot analysis was conducted to conclude whether converting big data into a geographical map could provide a precise geographical and spatial diagnosis regarding women's safety.

## 2. Policy Trends and Examples of Usage for Spatial Big Data

- “Geospatial data” refers to data that converts real-world geographical features and structures into rasters or vectors and provides access to relevant attributes.
- Spatial big data refers to the conversion of big data with location information into geospatial data, and can refer more generally to existing vector and raster-formed geographical and spatial information such as cadastral maps, road maps, DEM (Digital Elevation Models), three-dimensional imaging, etc.
- Construction of spatial big data projects are being carried out both domestically and abroad.
  - In conjunction with spatial big data construction projects in 2014, The Ministry of Land, Infrastructure and Transport conducted projects to develop technology for the analysis and utilization of spatial big data, providing big data platforms and services for the analysis and projection of universally salient events such as crime, natural and man-made disasters.

### 3. Results of a Pilot Study on Spatial Big Data for Women's Safety

- Our pilot study on the use of spatial big data for women's safety involved an analysis of the correlation between risk ratings for violent crime and spatial characteristics in the Gangnam-gu area of Seoul.
  - Empirical analysis was conducted on the effects that spatial characteristics (independent variable) such as transient population, credit card sales at nightlife businesses, police stations, video surveillance cameras, and security had on the risk rating of sexual violence (dependent variable) within the area in question.
- Data for analysis was formulated to examine how numerous variables affected the level of sexual violence occurring in the area and analyze spatial correlation.
- The table for analysis is defined as the XY coordinates of sexual violence occurrence, average credit card expenditures within a 100-meter radius of each location by risk rating, distance to the nearest police station, distance to the nearest surveillance camera, distance to the nearest safety lights, and the average transient populations of women in their 20s and men in their 30s within the 100-meter radius.
  - Not taking spatial correlation into account, the OLS of credit card expenditures, distance to the nearest police station, distance to the nearest safety lights, and the female and male transient populations all converted to meaningful results.
  - Among those factors, the average male population in their 30s and female population in their 20s have a positive (+) correlation at

a significant level of 0.01; this indicates that a higher population of males in their 30s and females in their 20s results in a higher risk of sexual violence.

#### 4. Methods of Utilizing Spatial Big Data for the Women's Safety Sector

- Geospatial methods can provide precise diagnosis and solutions specific to the problem.
  - Geospatial approaches have been found to be useful for developing customized policies that are effective in the field of women's safety.
  - Spatial big data analysis must be able to simulate policy effects to determine which will be most productive before final selection and implementation of policy alternatives.
- Geospatial methods can enable the production, publication and integration of chronologically and spatially accurate crime data
- We recognize the need to construct a shared platform and interactive real-time map for women's safety.

### V. Big Data-based Analysis of Female High-risk Drinking and Policy Alternatives

#### 1. Background and Purpose of Research

- A. The drinking rate of Korean women, especially high-risk drinking rate, is constantly increasing, yet there is limited information on

the characteristics of women's drinking due to a lack of existing research methods and subjects on the topic.

- Existing research is insufficient since it comprises primarily of research based on self-reported data, and thus lacks analysis based on objective data (distributed data)
- Research on female drinkers lacks attention to contextual factors and skews towards analysis of compositional factors

B. In this study, we intend to use big data to analyze the health effects of women's alcohol consumption and the use of medical services related to high-risk drinking, in order to suggest policy alternatives for prevention of high-risk drinking among women and the resultant harmful effects.

## 2. Study on the Current State and Trends of Female High-risk Drinking

- A. The number of high-risk female drinkers identified by the National Health and Nutrition Survey has continuously increased, showing that rates for men and women are converging.
- Monthly drinking rate, monthly binge drinking rate, and high-risk drinking rate are all increasing in women at faster rates than in men, with the high-risk drinking rate of women in particular increasing at a rate that is reducing the overall disparity between genders.
- The high-risk drinking rates of women in their 20s and 30s is increasing particularly sharply.

B. The high-risk drinking rate of Korean women varies by age and income level.

- For women, lower income levels correlate to greater high-risk drinking rates, but for men, the higher the income, the greater the high-risk drinking rate. For women in particular, the lower the income, the higher the rates of monthly binge drinking and high-risk drinking.

### 3. Study on the Current State and Trends of Female Liquor Consumption Using Credit Card Big Data

A. Credit card big data was analyzed using the data of Shinhan Card Co. customers aged 19 or older with either a credit or debit card using credit cards in businesses that primarily sell alcoholic drinks. The analysis includes the number of such alcohol consumers using credit cards, the number of alcohol consumption cases, and the cost of alcohol consumption over the past five years (2013-2017).

- Alcohol consumption by women has increased by about 10 percentage points since 2013.
- In particular, alcohol consumption among women in their 20s and 30s was higher than that of other age groups, which is consistent with the rapid increase in high-risk drinking rates among women in their 20s and 30s, as revealed in the analysis of National Health and Nutrition Survey data.

B. Alcohol consumption using credit cards is evidently linked to associated activities.

- For both men and women, the more that an individual used a credit card for alcohol consumption, the more likely they had used credit cards for karaoke rooms, convenience stores, and leisure sports.
- Furthermore, a higher amount of alcohol purchases was found to correlate to a higher frequency of visiting hospitals and oriental medicine clinics.

#### 4. Study on the Effects of Female High-risk Drinking Using Health Insurance Industry Big Data

- A. Using the big data of the National Health Insurance Service, we analyzed the number of cases of using medical services due to drinking and the resulting medical expenses over five years from 2013 to 2017.
- B. The rate of use of medical services for drinking-related diseases of those with different drinking levels differed by gender and age.
- The ratio of the number of people using medical services and expenses due to drinking-related health issues to the total number of people using any medical services and total expenses rose as for both men and women with the increase of age, with the highest ratios in their 50s and 40s. We speculated that the high-risk drinking rate and alcohol consumption shown in the analysis of National Health and Nutrition Survey data and credit card big data didn't result in an immediately apparent increases of using medical service, but did over a longer period of time.
- We can deduce that high-risk drinking among low-income populations may lead to poorer health conditions as the ratio of the number of

medical services and expenses for alcohol-related issues increased amongst low-income individuals.

## 5. Suggested Policies on Female High-risk Drinking

- A. The effects of consumer-based commerce on the increase of women's high-risk drinking should be considered.
  - Marketing and advertising by liquor companies, impact of mass media, release of lower-degree alcoholic products, etc.
- B. Policies to reduce the number of drinking disorders must be tailored to specific ages and social strata.
  - Given the external effects of female drinking (childbirth and childcare), policies for women in their 30s who are low income or are employed in sales services should be prioritized.
- C. It is noteworthy that general statistics and big data analysis have returned consistent results on the phenomenon of high-risk drinking among women.

## VI. The Apparent Impact of the “MeToo” Discourse in Social Big Data

### 1. Necessity and Purpose of Research

- As much of the “MeToo” movement's propulsion was fueled by social media's unique ability to propagate, social networks and the

“MeToo” movement are closely related. The “MeToo” movement of Korean society was also being waged over social networks and media. The trajectory of the “MeToo” movement on social networks is complex, with resistance being raised over time after initial unilateral approval and support, as well as instances of re-victimization. This study analyzes how the “MeToo” discourse has evolved in online social media spaces.

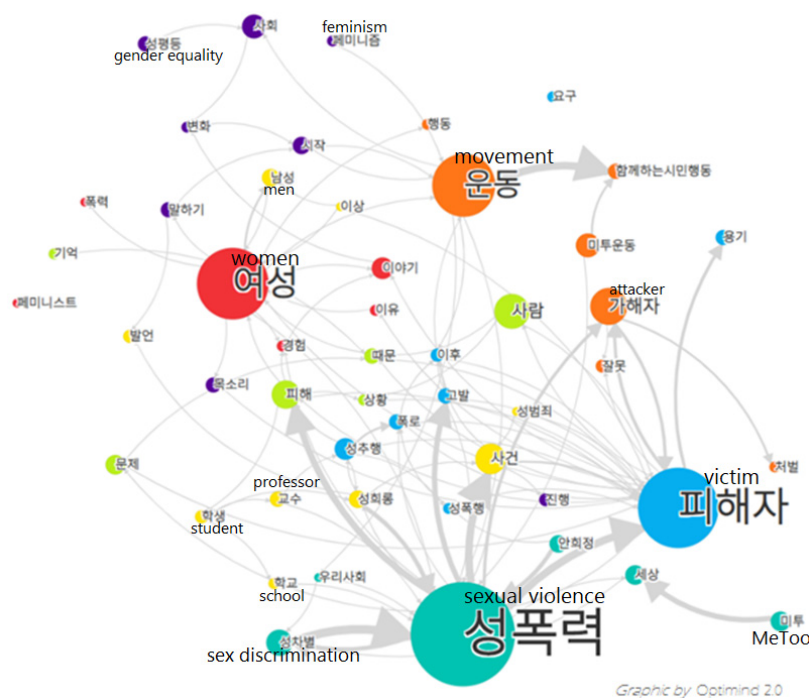
- The data collection and basic analysis of this study was commissioned by big data analysis company Ars Praxia (CEO Kim Do-hoon). The subject of analysis is Facebook's public postings, and the period of data collection is from Jan. to Aug., 2018. The search term was limited to “MeToo,” and original posting and comments which mentioned “MeToo” in the text of Facebook's public postings were collected. To select the texts to analyze, ‘potential pageview indicator’ was calculated and applied to filter posts with higher ripple power.

## 2. Key Research Results

- The analysis of postings revealed the number of public posts collected in the past eight months totaled 2,415 with a daily average number of 10.41 and monthly 301.88. The monthly distribution shows a sharp increase in March, but has decreased since April to less than 200 per month.
- By analyzing the contents of posts, shares, and comments, it is shown that people focuses on “revealing” or “accusing” the victimization of “women” due to “sexual violence,” “sexual harassment,” and “sex discrimination” rather than on fundamental reflections on gender issue. Moreover, due to Facebook's characteristics, the weight and influence

of “MeToo” in schools is more prominent than that of the workplace.

- There are many online spaces for people in their teens to 20s, including the “Bamboo Forest” (an anonymous bulletin board) at each university, so there are many forums for public discussion of issues related to school. We speculate that the fact that Facebook provides few appropriate spaces for public discussion of workplace issues, and that anonymity is not guaranteed, hinders an active online workplace “MeToo” movement in Facebook.

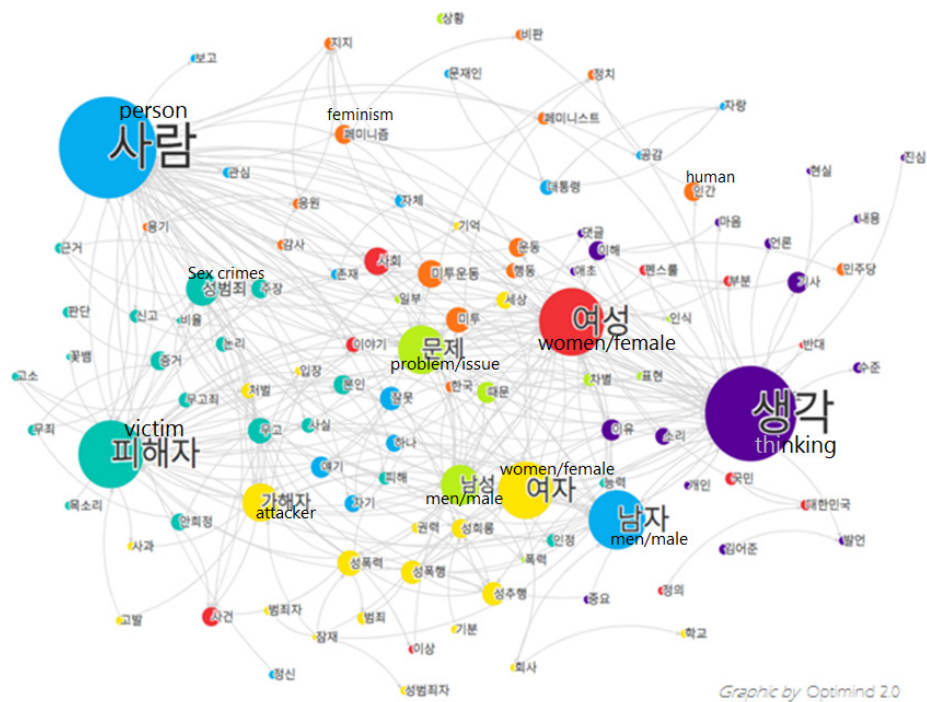


Source: Social Data for the Analysis of Spreading “MeToo” Discourse

[Figure 2] Content Analysis of the Concepts in Postings

- Our analysis of shares and comments indicates that, in contrast to the original postings, the comments reflect various conflicting

opinions and adversarial positions regarding “MeToo.” Furthermore, discrimination and reverse discrimination are widely discussed, with the opinions that women are not the only possible victims and men can be directly or indirectly victimized by the movement.



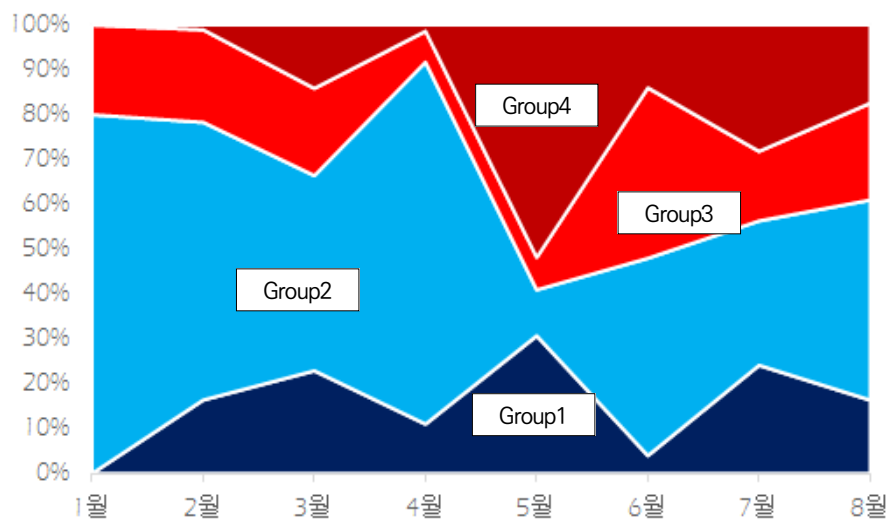
Source: Social Data for the Analysis of Spreading “MeToo” Discourse

[Figure 3] Content Analysis of Comments, Shares, and Replies

- Influencer analysis results: 1,000 influential users were derived based on the score of the ripple effects for each social action by user. We analyzed the characteristics of their social action by categorizing them into five groups based on their propensity. Group 1 is a “MeToo” proponent group that uses hate speech (radical support group), Group 2 is a proponent group that does not use hate speech (moderate support group), Group 3 is a reserved-position group that

does not use hate speech (moderated reserved group), and Group 4 is a reserved or opposing group that uses hate speech (radical opposing group). Group 5 is a group that cannot be judged regarding this topic or those whose posts has nothing to do with the topic.

- Regarding the distribution patterns of the comments and replies of influencers, the proportion of Group 2 (moderate support) decreases after April, and the proportion of Group 4 (radical opposition) tends to increase.



Footnote: Excludes Group 5

Source: Social Data for the Analysis of Spreading “MeToo” Discourse

[Figure 4] Proportion of Different Influencer Groups’ Comment and Re-comments by Month

### 3. Implications of Policies related to the “MeToo” Movement

- In order for the “MeToo” movement to succeed, efforts are needed

to resolve this structure of opposition. Extended conflict is not only wasteful, but could also become violent as antagonism continues to deepen. A practical alternative could be to convert the movement into the frame of perpetrator and victim of unequal power relations rather than that of male-female confrontation.

- The “MeToo” movement is also a sudden eruption of the cases of sexual violence that have accumulated over time, as the existing judicial system failed to provide sufficient protection against sexual violence. Legal fairness needs to be ensured along with proper social policies. Many discourses regarding Facebook's “MeToo” movement emphasizes judicial judgment.

## VII. Proposed Policies for the Production and Utilization of Big Data

### 1. Establishing Production Systems of Big Data for Women and the Family

A. Producing a specific scope of big data on women and the family

- The scope of usage of big data in women and family research varies considerably. A minimum boundary of big data for women and family research needs to be established.

B. Improving the use of big data in the study of women and the family

- The purpose of the pilot analysis is to identify the issues in the usage of big data. The results show that big data allows for the

exploration of new kinds of research topics that were not accessible through traditional statistics data, including subconsciously rooted issues, changes in detailed human behavior that differ over time and space, and the relationship between space and facility and human behavior.

- However, there are many challenges in collecting and utilizing data for a pilot analysis. In spite of the fact that the required data is not classified as personal information, the specific data was not available to the public. Therefore, we had to adopt proxy variables, which consequently reduced the level of analytical power. In the future, more reliable big data research will be possible only if research institutions establish cooperative relationships with the governmental organizations managing such information, or if they obtain necessary data through collaborative research.
  - If personal information of two or more types is linked, another level of analysis would be possible. But such activity is prevented by privacy protection policies, which resulted in the replacement by group variables. It is necessary to develop technology such as blockchains so that users can actively utilize data while protecting the data provider. Furthermore, the Personal Information Protection Act should to be revised to better facilitate policy research.
- C. Development of research methodology suitable for big data analysis
- Existing research on women focuses on gender differences. Most statistical analyses used a causal analysis framework to identify the differences in dependent variables caused by some demographic and socioeconomic independent variables.

- However, social network data frequently used in big data research often does not provide identification of individual characteristics. Thus there is a need to develop suitable research methodologies for using this type of data.

## 2. Measures to Prevent Discrimination in Big Data

- Big data algorithms run a significant risk of potentially reproducing and reinforcing various pre-existing social prejudices, including gender discrimination. Despite these risks being ingrained into the entire life process of big data, there is currently no effective system for testing or monitoring against these deficiencies.
  - In particular, algorithms that are involved in making decisions regarding individual valuations and eligibility present a risk of discrimination that can have profound repercussions. Therefore, we should make an effort to fully understand this potential for discrimination and make it a point of concern in the process of developing and utilizing algorithms.
- A. Facilitating research on the reproduction of prejudice in big data algorithms
- We recommend that the government promote research on whether big data algorithms are reproducing prejudices when involved in decision-making on individual eligibility, especially those in the financial and professional fields. Big data algorithms in the early stages of development present an opportunity to promote fairness while contributing to the development of related industries and human resources practices.

## B. Composition and proposition of guidelines

- It is necessary to present guidelines that can help prevent the potential issues and problems that are present throughout the development of algorithms. In order for these guidelines to be practically effective, they must be specific and specialized to a particular type of big data and the development method and utilization field of the algorithm in question.

## C. Development and circulation of algorithm prototypes

- The National Statistical Office operates the National Accreditation Statistics System and the Statistical Quality Assessment System to enhance the reliability of national statistics. Although no government department is currently tasked with overseeing the quality of big data, it is necessary to develop and circulate prototypes of algorithmic meta-data in order to promote the accountability and transparency of big data algorithms.

## D. Establishing a specialized center for gender-discriminatory algorithms

- As big data algorithms are being universalized in all fields of industry, the influence of algorithms in our lives is gradually becoming momentous. However, if we rely only on the autonomous efforts of algorithm developers answering to market logic and institutional objectives, it would be difficult to guarantee the accuracy and equity of algorithms. Therefore, it is also necessary to develop service industries that can keep these forces in check by auditing algorithms independently of the organizations that develop them.

- With regards to algorithms with gender biases, empirical research must be conducted on the prevalence of algorithmic gender discrimination in areas that are directly pertinent to individual opportunity and quality of life, such as credit rating, loan screening, insurance review, job recruitment, and school admissions. Likewise, it is necessary to study which issues in the life cycle of big data result in such gender discrimination. This information can directly contribute to developing practical guidelines for the prevention of gender discrimination by algorithms.
- In addition, we suggest the establishment of a specialized center within public research institutes dedicated to addressing algorithmic gender prejudices by receiving complaints of alleged discrimination and conducting audits on suspected problematic algorithms.

## Appendix 1.

National Research Council for Economics, Humanities and Social Studies  
Cooperative Research Series

“Improve and discover policies through the use of big data by women's families (II)”

### 1. Series in Cooperative Research

	Research Report Titles	Institutions
18-56-01	Improvement and Discovery of Policies through the Use of Big Data on Women and Family (II)	Korean Women's Development Institute
18-56-02	High-risk Drinking Analysis and Policy Alternatives of Women Using Big Data	Inje Univ. Industry Academy Cooperation Foundation
18-56-03	A Study on the Utilization of Space Big Data in the Women's Safety Sector	Korea Research Institute for Human Settlements

## 2. Participants

Research Institutions		Chief of Research	Participating Researchers
Chief Research Institution	Korean Women's Development Institute	You Kyung Moon (Senior Research Fellow)	<ul style="list-style-type: none"> <li>• Young Taek Kim (KWDI Research Fellow)</li> <li>• Ki Taek Jeon (KWDI Research Fellow)</li> <li>• Sungmi Jung (KWDI Associate Research Fellow)</li> <li>• Ho Joong Bae (KWDI Researcher)</li> <li>• Hee-Tae Chung (KWDI Visiting Researcher)</li> <li>• Youjing Kim (Ars Praxia, Research Fellow)</li> </ul>
Contract Research Institution	Inje Univ. Industry Academy Cooperation Foundation	Kwang Kee Kim (Professor, Inje Univ. Graduate School of Public Health)	<ul style="list-style-type: none"> <li>• Jung JeKarl (Prof. Ewha Univ. Dept. of Health Convergence)</li> <li>• Taemin Song (Prof. Samyuk Univ. Dept. of Health &amp; Human Performance)</li> <li>• Minjoo Choi (Doctoral student, Inje Univ. Graduate School of Public Health)</li> </ul>
	Korea Research Institute for Human Settlements	Daejong Kim (Research Fellow, Korea Research Institute for Human Settlements)	<ul style="list-style-type: none"> <li>• Ryunghyuk Im (KRIHS Research Fellow)</li> <li>• Gaweon Shin (KRIHS Research Fellow)</li> </ul>



**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>