

---

# Introduction

# Quantitative Methodologies and Big Data

BY INGE HENNINGSSEN, TINNE STEFFENSEN  
AND HILDA RØMER CHRISTENSEN

---

‘Data has become a currency of power. The most successful Internet businesses make their money by aggregating data. Decisions of public import, ranging from which products to market, to which prisoners to parole, to which city buildings to inspect, are increasingly being made by automated systems sifting through large amounts of data.’

(Data visualization from a feminist perspective, Interview with Catherine D’Ignazio, this volume)

**T**his volume of *Women, Gender & Research* addresses the emerging interest in quantitative methodologies and big data in women’s and gender studies in the global North and West. It reflects the growing hegemony of evidence-based views in neoliberal policy-making, which has turned statistics and quantitative methodologies into key data with wide-

ranging effects at both the institutional and individual levels. As a result, knowing how to collect, find, analyze and communicate data is of increasing importance in present-day society. This development has made ownership of data pivotal, along with access to IT equipment, resources and expertise. Data is, as argued by several authors of this issue, today mainly collected and stored by big corporations and governments, who have the resources to do so and who often control access. People today, as argued by Catherine D'Ignazio, are far more likely to be discriminated against with data or surveilled with data than they are to use data for their own civic ends.

What is more, this volume of *Women, Gender & Research* aims to explore critical aspects of power and inequality in prevailing quantitative methodologies and big data, as well as to accentuate the potentials of alternative or even subversive uses of 'big data' and new technologies of collection and visualization. How does the provision of new data feed into the practices and politics of social and gender equality? Is it possible to collect and organize data collections in ways that support new forms of democratic governance? And what are the potentials and pitfalls of emerging methodologies? How can bodies be made visible without creating new essentializing categories?

While addressing current issues, this volume also seeks to nuance or even overcome the old 'paradigm war' in feminist scholarship between quantitative and qualitative methodologies. It is well known that gender research took off in the 1970s along with political currents such as the student revolt and the women's liberation movements. This also implied a break from established and dominant scientific paradigms, and not least from quantitative methodologies in the social sciences. In parallel with other oppositional trends, feminist research defined itself as opposed to a 'positivist, quantitative research methodology', which became regarded as the bias of mas-

culine knowledge and women's invisibility. Feminist criticism of quantitative research culminated over the first decades of women's and gender studies in the 1970s, being concerned with the power implications of research methodologies. It was argued that engaging with quantitative research implicitly supported sexist values on a broad scale. It was further argued that female subjects were excluded and marginalized and that the relationship between researcher and research subjects was intrinsically exploitative. Moreover, the resulting data were regarded as superficial and overgeneralized, and it was argued that quantitative research was not being used to overcome social problems (Oakley 1998: 709). In contrast, feminist research was branded as research with, for and about women. And appropriate methods included participant observation, semi-structured interviewing, life-histories and focus groups. Such methodologies came to be seen as epistemologically distinct from the quantitative methods of surveys, experiments, statistical records, structured observations and content analysis. All in all, feminist scholarship contributed to the paradigm debate by introducing new themes which confronted the gender-blind and sexist core of much research (Oakley 1998: 708).

In the twenty-first century, several scholars have entered the field of quantitative methodologies and seem to be in favor of ending the war and bridging the debate (McCall 2005; Hughes & Cohen 2012). Two alternatives seem to have come to prominence. One is seen in the efforts to develop research practices where quantitative and qualitative methodologies are applied in a mixed-methods approach. Another alternative is to accept the academic division of labor and specialization and the fact that particular methods can have particular assets and limitations, issues that are related to research questions and research interests (McCall 2005: 1791).

In 2005 the American sociologist Leslie

McCall wrote a path-breaking article in which she introduced fresh ideas and bridge-building efforts in the field of feminist research and quantitative methodologies. In so doing McCall addressed the more theoretical idea of intersectionality as introduced by Kimberle Crenshaw (1989) and connected it to a broader discussion between epistemology and ontology. Do intersections take place between already fixed categories, such as gender, ethnicity and class, or do they produce new categories? McCall suggested a systematic division of the complexity of intersectionality into anti-categorical, intra-categorical and inter-categorical complexities. Here inter-categorical complexity is used by researchers who adopt existing analytical categories to document relationships of inequality among social groups along multiple and conflicting dimensions, while intra- and anti-categorical complexity were seen as being connected to qualitative micro- and meso-level studies.

In line with McCall's argument, the articles of this volume – coming from within the interdisciplinary field of gender studies – demonstrate that quantitative methodologies and big data by no means represent a unified and fixed field in the context of gender studies.

The contested role of quantitative methodology overall in feminist research is introduced through a theoretical discussion of essentialism and deconstruction in feminist theory. In the article *Reconciling anti-essentialism and quantitative methodology* by Mathias Fjællegaard Jensen, it is argued that the essentializing implications of quantitative methodology might prove less problematic if one keeps a strategic or political feminist aim central in a given research project. In so doing, Jensen considers a range of central concepts, such as Irigaray's notion of mimicry, Spivak's strategic essentialism and Butler's contingent foundation. Such theoretical deliberations are then connected to a specific use of ideas of variables in quantitative analysis, through which so-

cial categories can be deconstructed quantitatively. Quantitative deconstruction as a methodological approach could potentially enrich both the theoretical and empirical understandings of the variables in question and of the underlying identities of the researched subjects.

## MULTIDIMENSIONALITY IN VARIOUS FORMS

In the article *Intersectionality: an inter-categorical empirical approach*, Ruth Emerek demonstrates how basic empirical quantitative methods help reveal intersectional complexity, which may enlighten both quantitative and qualitative research. This is done by means of three examples based on statistical data from Denmark. First, it is shown how a narrow focus on the gender category without its intersections may produce misleading results. Secondly, it is demonstrated that an overly narrow focus on the intersection of two categories (gender and educational attainment) may hide the overall effect of the categories. The last example, focusing on pay gaps, shows how intersectional effects are revealed by comparing the results of an additive multiple analysis with the results of a separate analysis for women and men in the private and public sectors. By applying an inter-categorical approach and large data sets, the article demonstrates how it is possible to incorporate multiple categories in descriptions of gender inequality, differences and similarities, and to investigate if intersections should be included.

The following articles represent relatively recent developments in quantitative methods that seek to counter the causal, general linear modelling that has dominated quantitative methods for so long. They both argue for the use of descriptive methods that highlight the possible multidimensionality, pluralisation and heterogeneity of data.

According to Claus D. Hansen, traditional sex-difference research has for too long been too simplistic in its comparison

of men and women on different parameters, while ignoring the possible heterogeneities inherent in the social categories of ‘men’ and ‘women’. In the article *An alternative approach to the analysis of gender differences: geometric data analysis*, Hansen argues that many of the problematic features of traditional statistics can be solved by using geometric data analysis, as this makes possible multiplicity, individual-level analysis and visualization. The value of this method is demonstrated using a case study of survey data from Danish vocational schools, where students were asked a range of questions about their educational attributes. In spite of the students on average showing gendered differences in job attributes, the patterns become more complex when more components were included. Inspired by McCall’s intra-categorical approach, the significance of educational institutions is also explained.

Following this, Tinne Steffensen addresses anew the century-old interest in demographic changes in family formation and parenthood. In the article *The roads more or less traveled: a sequence analysis of family formation and parenthood for a cohort of Danish women born in the 1970s*, Steffensen departs from the increased attention to when and how many times Danish women give birth. The concern signified by this increased attention derives from an entanglement of low fertility and increased age at first birth, along with the development of assisted reproduction technologies. Steffensen demonstrates the potential of sequence analyses of family formation by analyzing a randomized sample of 1,500 women born in 1973 and 1974. Through sequence analysis of longitudinal registry data, she identifies seven distinct clusters (i.e. typologies) of family formations in Denmark. The study thereby confirms that the first child is a constituting factor of the nuclear family, which often precedes marriage. However, the identified clusters also show great variation when it comes to age at birth of the

first child, region, socio-economic status and overall turbulence in their trajectories.

#### ABSENT BODIES, POWERFUL DATA, AND VISUALIZATIONS

In Nanna Thylstrup and Kristin Veel’s interview with Catherine D’Ignazio, *Data visualisation from a feminist perspective*, D’Ignazio introduces readers to a practical feminist approach to the power and political implications of collecting, analysing and disseminating data visualization. According to D’Ignazio, the main problem with data visualization remains the ‘missing body problem’, where bodies are extracted, absent, uncounted and rendered invisible from the data presentation. To counter this problem, D’Ignazio suggests six design principles for the content, form and process of discovery that can help foster feminist data visualization. Among these are design justice, co-design, and participatory design.

The following three articles provide different angles on the issue of data and visualisation; through an inquiry of open data collections and its consequences, absent bodies in online data visualisation and an example of how the participatory design might be useful in developing gender responsive indicators.

Open-data collections can be powerful, providing democratic tools to illustrate women’s health across Europe, as argued in the article *Determinants of women’s health in Europe: using large open-data collections to unveil the hidden part of the iceberg*, by Lourdes Cantarero-Arévalo. The article describes benefits offered by the large volume of open-access data (e.g. from WHO, Eurostat and OECD), in comparison to access-restricted big data. Besides an overview of the main publicly available databases which gather sex-disaggregated data information, the article presents their strengths and limitations. Open online data collections can be used as tools to argue in favour not only of the implementation of health-care policies,

but also of social and economic policies aimed at improving women's health in Europe. Yet open-data collections need continuous monitoring and updating to ensure reliability of data from all countries around the globe, and at the same time need to guarantee individuals' anonymity.

In *The political potential of numbers: data visualisation in the abortion debate*, Rosemary Hill uses Google Image scraper to explore how the anti-abortion agenda dominates online data visualizations of abortion, while these visualizations also decontextualize abortion from women's lived and bodily experiences. This leads her to argue that it is vital for feminists to work with data visualization in order to critically counter and challenge the idea that data visualization carries the potential and power to change the world (for the better).

The final article of this volume deals with how to change the production and collection of data for the better. *The development of gender-responsive indicators: towards a participatory approach*, Michèle Amacker, Isabelle Schlaepfer, Christine Bigler and Andrea Graf argue that, although much attention has been paid to the size and possibilities of big data, including in the field of gender equality, there has been too little concern with the quality of the indicators being measured. The authors therefore suggest a participatory research design where important stakeholders and target groups and their social contexts actively participate in the development of gender-responsive indicators for measuring gender equality.

## REFERENCES

- Crenshaw, Kimberle (1991): Mapping the margins: Intersectionality, identity politics, and violence against women of color, in: *Stanford law review* 1991/42(6): 1241-1299.
- Hughes, Christina & Cohen, Rachel Lara (eds.) (2012): *Feminism counts: quantitative methods and researching gender*. Routledge, Abingdon.
- McCall, Leslie (2005): The complexity of intersectionality, in: *Signs* 2005/30(3): 1771-1800.
- Oakley, Ann (1998): Gender, methodology and people's ways of knowing: some problems with feminism and the paradigm debate in social science, in: *Sociology* 1998/32(4): 707-731.

Inge Henningsen has a degree in statistics and is Emeritus Associate Professor at the Institute of Mathematics at University of Copenhagen. She has been a member of the research project *Gender Barriers in Advanced Studies and Research* 1996-2002 and has published extensively employing quantitative methods in gender research.

Tinne Steffensen is MSc in Sociology from University of Copenhagen and is currently working at the Center for Public Innovation as a data analyst. Her areas of interest are gender and fertility, data analysis of longitudinal data, health and the public sector.

Hilda Rømer Christensen is Associate Professor and Head of the Co-ordination for Gender Studies, University of Copenhagen. She also leads the Gender Certificate, an interdisciplinary educational initiative at the University of Copenhagen, where she, among other things, is course coordinator on the summer school *Gendering quantitative methodologies*.