sh!ft

THE PROJECT TO END DOMESTIC VIOLENCE

Using Data Science to Engage and Mobilize Men for Violence Prevention and the Advancement of Gender Equality, Diversity, Justice, and Inclusion: Rapid Evidence Review

2022

UNIVERSITY OF CALGARY
FACULTY OF SOCIAL WORK

Authors

Elizabeth (Liz) Baker
Laura Pascoe
Lana Wells
Naomi Phung

Acknowledgements

Shift acknowledges the support of Women and Gender Equality Canada, which made this valuable project possible. The authors would also like to thank Winta Ghidei for sharing her guidance and knowledge on rapid research reviews, and Elena Esina and Winta Abera for their project and organizational support. Finally, we would like to thank the men who are leading this work in Canada and who took the time to contribute their expertise to advance this important work.

The views of the authors do not necessarily represent the views of Women and Gender Equality Canada or the Government of Canada.

Suggested Citation

Baker, E., Pascoe, L., Wells, L., & Phung, N. (2022). *Using data science to engage and mobilize men for violence prevention and the advancement of gender equality, diversity, justice, and inclusion: Rapid evidence review.* Calgary, AB. The University of Calgary, Shift: The Project to End Domestic Violence.

Contact

Lana Wells, Brenda Strafford Chair in the Prevention of Domestic Violence
Faculty of Social Work, University of Calgary
2500 University Drive NW, Calgary, AB, Canada T2N 1N4
Phone: 403-220-6484 Email: lmwells@ucalgary.ca

Table of Contents

Author's Note **2**
Executive Summary **3**
1.0 Introduction **5**
2.0 Methods **6**
3.0 Results **8**
  3.1 Source characteristics **8**
  3.2 Figure 1. PRISMA flowchart of study selection process **9**
  3.3 Figure 2. Summary of studies reviewed **10**
4.0 Findings: How are data science methods being used to promote social change? **13**
  4.1 Creating accurate characterizations of society **13**
  4.2 Identifying inequity and/or promoting social justice **14**
5.0 Findings: What are the key data science methods? **15**
  5.1 Text-based methods **15**
    Figure 3. Example of semantic network analysis visualization 18
    Figure 4. Example of cluster analysis visualization 19
  5.2 Predictive modeling **19**
6.0 Findings: What are the key considerations and cautions? **21**
  6.1 Data fundamentalism **22**
    6.1.1 Data sets 22
    6.1.2 Analyses 24
    6.1.3 Privacy 25
7.0 Recommendations: What are key considerations and cautions for those in the field of engaging men? **26**
    7.0.1 Awareness and education 26
    7.0.2 Consider how data science can be used as a tool to mobilize men 26
    7.0.3 Increase and improve data collection 27
    7.0.4 Create safety nets 27
8.0 Conclusion **28**
References **29**

Author's Note

We, the authors, would like to take this opportunity to situate ourselves in relation to this research and flag some of the tensions that we continue to navigate as feminists working to advance gender and social justice. First, we are white settlers, trained in the Western scientific tradition, with extensive experience working with feminist issues from an intersectional perspective. Each of us has extensive experience working directly with men in the areas of violence prevention and gender equality. Based on our experience, we firmly believe that gender and social inequality is inextricably linked with rates of male violence against all genders and our interventions must focus on all forms of violence to stop violence before it starts.

We are also white feminists committed to advancing racial justice and are on an ongoing journey to understand and learn more about where and how we can be most useful in this work. At Shift, we have been integrating approaches that aim to call *in* rather than *out*, while also reflecting on our own practices and building creative and innovative skills, so that we can maximize our capacity to hold people accountable in ways that generate healing, recovery, repair, and prosocial change. We believe it is imperative to ask hard questions and think strategically about what is and is not working in efforts to achieve social change across anti-violence, gender equality, and justice, diversity, and inclusion fields so that we can build momentum for bigger and more impactful movements.

In completing this review, our methods and analysis used an intersectional approach which allowed us to clearly see the dearth of research on strategies to engage and mobilize men at the intersections of gender equality, violence prevention, and advancing equity, diversity, justice, and inclusion. We worked diligently to name and map the ways in which these gaps need to be addressed, but we recognize that our analysis may have shortcomings as we continue the process of learning and unlearning in relation to our own positionality and context in this work. We welcome those who want to call us *in* so that we may continue to make our work stronger, more relevant, and more impactful across a wider audience.

In solidarity,
Liz, Laura, Lana, and Elena

Executive Summary

CallinMen: Mobilizing More Men for Violence Prevention and Gender Equality in Canada is a knowledge synthesis research project led by Shift: The Project to End Domestic Violence, a primary research hub with the goal to stop violence before it starts. Shift is based out of the Faculty of Social Work at the University of Calgary (Shift/UCalgary). As part of the *CallinMen* project, nine rapid evidence reviews were conducted on evidence-informed *primary prevention* approaches to engage and mobilize men to prevent and disrupt violence and inequalities, with the goal to share these findings with those funding and working with men and male-identified people to prevent violence and advance equity. To support and advance work to engage and mobilize men, both well-known and emergent approaches that show promise in engaging and mobilizing men were identified for review. Since there is no research available (to our knowledge) for how methods in data science have been utilized to engage men in violence prevention, this document reports on the findings for how methods in data science have been used to inform social change initiatives and how this evidence could be applied to engaging and mobilizing men to prevent violence and advance equity.

Our definition of data science: Data science is a multidisciplinary field of study that focuses on creating, collecting, handling, and analysing data in order to extract actionable insights from the large and ever-increasing volumes of data that are available across a wide range of platforms and sources. Data scientists use scientific methods, processes, algorithms, and systems to identify patterns and trends in data in order to identify issues, predict future events, reduce risk, and improve programming and outcomes.

How are data science methods being used to promote social change?
Five studies were included in this review, with one study taking place in Canada. Studies focused on gender bias in the media, inequities in education, inequities in rental housing decisions, the public's perception of sexual sports violence, and physical intimate partner violence during pregnancy and the postpartum period. However, none of the studies explicitly engaged men in relation to these issues. Two of the studies utilized data science methods for the purpose of portraying what was happening in society, while the remaining three utilized data science methods for the purpose of identifying inequities and/or promoting social justice.

The studies reviewed utilized different methods in data science. These were categorized as data science methods that a) collect and analyze text-based data from publicly available online sources to extract meaning, and b) use predictive modeling to predict risk and/or behaviour. It is important to note that a plethora of other data science methods exist, and the studies reviewed represent only a very small sampling of the methods available.

Insights from data science research
Data science holds incredible potential to identify patterns and trends in Big Data that can help inform which populations of men to target and how to engage and mobilize them most effectively. Two key ways to use data science in the field of engaging men are to:

1. Increase and improve data collection specific to the agenda of engaging and mobilizing men. In order to use data science methods, we need access to large data sets that are also of high

quality. One way to do this is to create and teach a consistent methodology for data collection that is done at scale.

2. Partner with those who already have in-depth knowledge and experience in using data science methods, and work to build capacity in the field of engaging men.

Cautions and considerations in using data science

- Without understanding the limitations of the data sets and/or the analyses, data science may be used incorrectly, perpetuate biases, present false information, and cause irreversible harm. For example, with machine learning, we do something called "training" which means we are teaching the computer to develop a certain model that can be used for prediction. To train, we have to utilize existing data. Unfortunately, the software/analyses/algorithms that are used to analyze the data will be biased if the systems and data (software/analyses/algorithms) they are trained on is biased (i.e., if a data set contains much more data from male participants as compared to female participants). Thus, we need to create data sets that are free from bias (i.e., no differences in representation between males and females) and train the software/analyses/algorithms on the unbiased data.
- The Data Equity Framework created by We All Count[1] is a useful tool to help people understand how each and every decision that is made with data (from developing the research question and sampling to analyzing and interpreting) is influenced by people's biases, and can therefore bias results.

1.0 Introduction

In 2020, Shift/UCalgary was awarded a research grant from Women and Gender Equality Canada (WAGE) for a knowledge synthesis research project entitled *CallinMen: Mobilizing More Men for Violence Prevention and Gender Equality in Canada*. Little knowledge synthesis work has been done to date to increase understanding of what strategies and approaches meaningfully engage and mobilize men to prevent violence and advance gender equality, diversity, justice, and inclusion in Canada; this research fills that gap. Specifically, *CallinMen* advances the state of knowledge by identifying and reviewing the evidence base for key strategies and approaches that show promise in engaging and mobilizing men to prevent violence and advance gender equality, diversity, justice, and inclusion in Canada, and develops an evidence-informed "behaviour change toolbox" that consolidates these strategies and approaches.

Therefore, to identify and review promising approaches to engaging and mobilizing men to prevent violence and advance gender equality, diversity, justice, and inclusion, nine rapid evidence reviews[i] of the academic and grey literature were conducted[ii] in 2021 with the goal to share these findings with those funding and working with men and male-identified people to prevent violence and advance equity. This document reports on the findings for how methods in data science have been used to inform social change initiatives. Since there is no research available (to our knowledge) for how methods in data science have been utilized to engage men in violence prevention, we summarize research focused on social change broadly. Following this, we critically analyze how these methods can be applied for engaging and mobilizing men to prevent violence and advance gender equality, diversity, justice, and inclusion in Canada.

In this review, we define data science as a multidisciplinary field of study that focuses on creating, collecting, handling, and analysing data in order extract actionable insights from the large and ever-increasing volumes of data that are available across a wide range of platforms and sources. Data scientists use scientific methods, processes, algorithms, and systems to identify patterns and trends in data in order to identify issues, predict future events, reduce risk, and improve programming and outcomes. We purposefully created a broad definition to be inclusive of literature, given the relatively nascent state of this research, especially as it applies to promoting social change.

It is important to note that this research project is focused on advancing *primary prevention* approaches, meaning that we are focused on identifying strategies that change the root causes which drive violence, discrimination, and gender inequality to prevent initial perpetration and victimization of violence, harassment, discrimination, and inequities[2]. In line with this focus, our research seeks to understand strategies and approaches that incubate and catalyze male-identified

---

[i] A rapid evidence reviews is a process that synthesizes knowledge through the steps of a systematic review, but components of the process are simplified or excluded in order to shorten the length of time required to complete the review. The process includes identifying specific research questions, searching for, accessing the most applicable and relevant sources of evidence, and synthesizing the evidence.
[ii] Rapid evidence reviews were conducted on: bystander approach, social norms approach, nudge approach, virtual reality, gamification, data science, fatherhood, calling in, and community justice.

prosocial behaviours and systems that prevent violence, harassment, discrimination, and inequality before they begin.

The specific research questions that guided the rapid evidence review were:

1. How are data science methods currently being used to prevent gender-based violence and/or advance gender equality, diversity, justice, and inclusion, particularly with regards to engaging men?
2. What are the key methods within the field of data science that are being used to inform social change initiatives?
3. What are the key considerations and cautions for using these data science methods?
4. For those in the field of engaging men interested in using data science to inform, design, and evaluate their work, what are key considerations and cautions in using data science?

2.0 Methods

A rapid evidence synthesis/review (RES) was conducted in August 2021. RES is "a form of knowledge synthesis that follows the systematic review process, but components of the process are simplified or omitted to produce information in a timely manner."[3] The process includes identifying specific research questions, searching for, and accessing the most applicable and relevant sources of evidence, and synthesizing the evidence.

A systematic search strategy was performed using a combination of keywords, limited to the title of the articles. The search performed was: ("data science" or "big data" or "data equity" or "equitable data" or "open data" or "open source data" or "Linked Data" or "data revolution" or "advanced analytics" or "prediction analytics" or "predictive analytics" or "predictive analysis" or "predictive modelling" or "prediction model" or "data mining" or "behavioral analytics" or "machine learning" or "artificial intelligence" or "Bayes Theorem" or "automation" or "algorithm" or "decision tree learning" or "natural language processing" or NLP) AND ("violence" or "gender-based violence" or "gender based violence" or GBV or "family violence" or "domestic violence" or "domestic abuse" or "intimate partner violence" or IPV or "dating violence" or "violence against women" or VAW or rape or "sexual assault" or "sexual violence" or "sexual abuse" or "sexual harassment" or "workplace harassment" or "sexual misconduct" or "consent" or "gender equality" or "gender equity" or "gender justice" or "gender parity" or "gender transformative" or bullying or empathy or belonging or "harm reduction" or prejudice or justice or diversity or equity or inequity or racis* or "anti-racis*" or antiracis* or Indigenous or "First Nations" or Inuit or Métis or "social good").

Searches were conducted in the following academic databases: EBSCO (All databases, including CINAHL, Academic Search Complete).

- EBSCO (All databases including CINAHL, Academic Search Complete)

Inclusion criteria:
Time frame: 2010-2021
Publication language: English.
Availability: Full text option only.
Literature had to meet the following inclusion criteria:

- Describes how data science is used to prevent gender-based violence and/or advance gender equality, diversity, justice, and/or inclusion, particularly with regards to engaging men, aged 18 and over
- Provides details on what data science methods are used and the process for how data science informs efforts to prevent gender-based violence and/or advance gender equality, diversity, justice, and/or inclusion, AND/OR
- Provides explanation and supporting evidence for issues, limitations, cautions, and/or key considerations in using data science for social good/social change
- Literature may come from anywhere in the world; however, priority will be to locate literature focused on Canada or in other countries with similar economic, social, and cultural similarities to Canada (such as the United States, Australia, New Zealand, England, Scotland, Wales, Northern Ireland, Republic of Northern Ireland, Norway, Denmark, Sweden, Finland and Iceland).
- Articles that *do not* meet the criteria but seem relevant/valuable will be included in discussion/recommendations or where appropriate.

Literature was removed based on the following exclusion criteria:

- Literatures does not provide clear description of how data science methods/techniques have been used for the purposes of preventing gender-based violence, and/or advance gender equality, diversity, justice, and/or inclusion, particularly with regards to engaging men
- Literature does not focus on primary prevention (e.g., responses to violence)
- Literature that focuses on populations under the age of 18
- Literature that is a commentary on other publication which is not included in review
- Literature is a protocol, and no publication on study findings can be located through an internet-based search.

Information was extracted in a standardized form, including the following: author, publication year, discipline (if available), topic(s), country, data science method, key findings (for empirical articles) OR implications (for commentaries), cautions/limitations, and key insights on using data science to engage and mobilize men (reviewer original input).

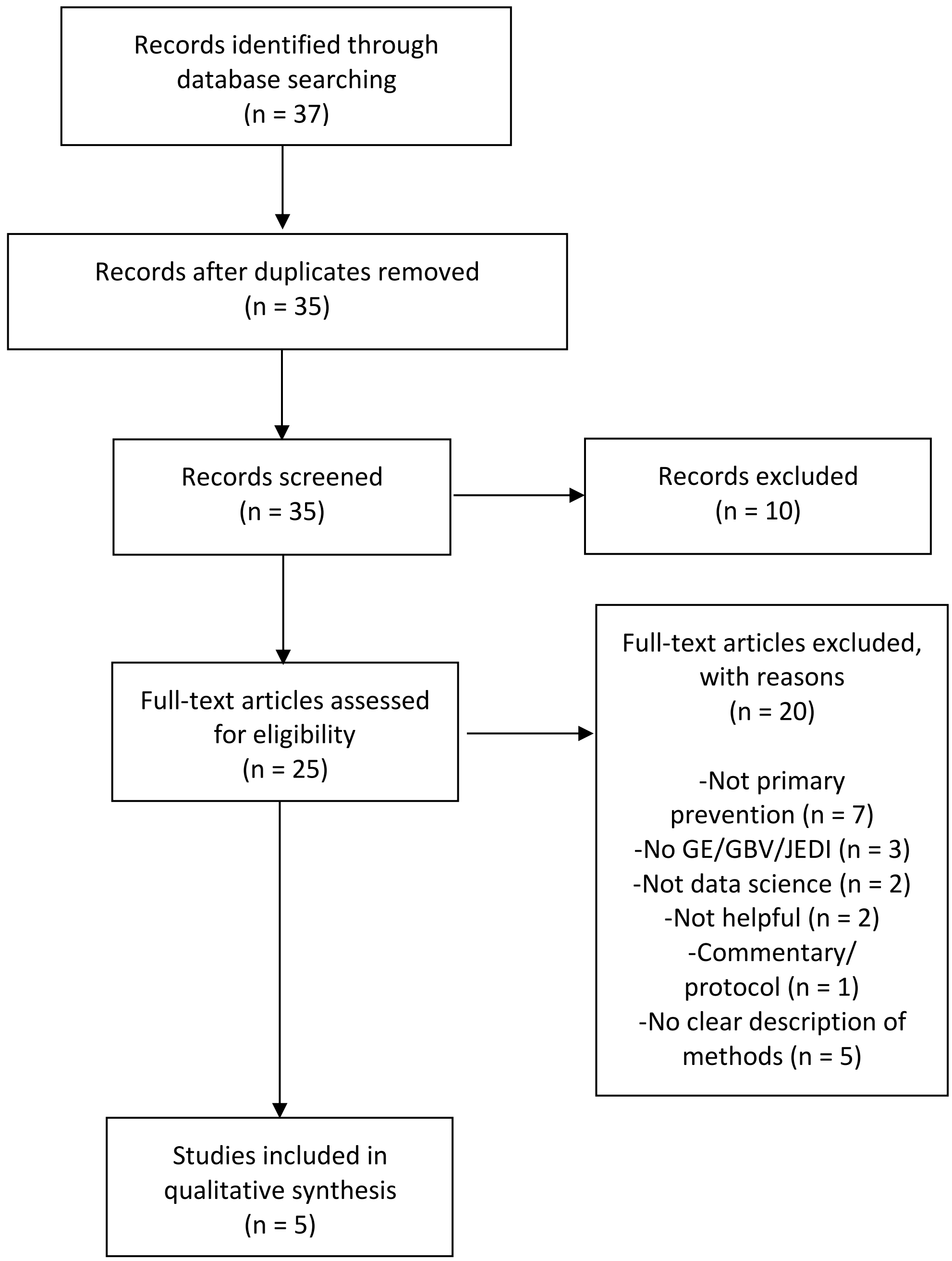
3.0 Results

3.1 Source characteristics

Across all databases, a total of approximately 400 search results were initially identified. Following screening, removal of duplicates, and full text analysis, five studies were included in the final assessment.

The five articles included in this review come from the following disciplines: educational psychology[4], linguistics[5], psychology/sociology/public health[6], epidemiology[7], and law[8]. The studies focused on inequities in learning/social justice in education[3], gender bias in the media[4], perceptions of sexual sports violence[5], physical intimate partner violence during pregnancy and the postpartum period[6], and inequities in rental housing decisions[7]. To be broadly inclusive, this review included both empirical articles[4, 5, 6] and commentaries[3, 7]. For empirical articles, detailed descriptions were given regarding the data science methods. For commentaries, descriptions were still provided, but in far less detail.

3.2 Figure 1. PRISMA flowchart of study selection process



3.3 Figure 2. Summary of studies reviewed

Authors	Type	Topic(s)	Purpose	Summary of Methods	Findings (empirical) OR Conclusions (commentary)	Cautions
Aguilar, 2018	Commentary	Inequities in learning; social justice in education	*Immediate purpose*: to identify students who are struggling, and then provide them with personalized learning tailored to their needs. *Long-term purpose*: to reduce social inequities in education and learning.	1. Data is collected from students. Often, this is in the form of administrative data and includes information such as student demographics, student grades, and students' behaviours (i.e., class attendance). 2. The data is analyzed to predict which students are most at-risk for various outcomes, such as low grades, dropping out, etc. 3. The data can be further analyzed to examine trends within and across students, courses, or populations. 4. Data can be displayed in data dashboards in real time to help both teachers and students understand these snapshots of information. 5. Learning analytics-driven technology then can offer suggestions for personalized learning tailored to the students' unique needs. This is done at-scale (i.e., across all students) and automatically (no teacher effort required).	Learning analytics can be used to identify which students are most at-risk of poor academic outcomes, and then provide them with personalized learning plans. This is critical, as learning environments typically teach to the "average student", meaning students who do not fit this profile often are put at a disadvantage. This methodology allows for personalized learning to happen at scale and automatically, which greatly reduces the burden of teachers.	There are concerns with privacy and consent. For example, some learning data dashboards automatically pull information and do not give students the opportunity to provide consent for this to happen. Even when consent is explicitly given at one point in time (beginning of the semester), it is debated if consent can hold for the duration of the course (throughout the semester) or beyond.
Asr et al., 2021	Empirical	Gender bias in the media	*Immediate purpose*: to provide information on how frequently women are quoted in the news as compared to men. *Long-term purpose*: to motivate news organizations to provide more gender diversity in their reporting.	1. Large-scale text processing (software called "Gender Gap Tracker") was used to pull text from news stories posted online. 2. The software extracted quotes from the data. 3. Natural language processing was used to identify the sources of the quotes and the authors of the article; then it identified who each quote belonged to. 4. Name entity recognition was used to assign gender to each source. 5. Errors were identified, and data was cleaned by humans. 6. Counts and percentages of female vs. male sources were counted overall, and for each news source. 7. Results were uploaded to a data dashboard.	Although more women were quoted as compared to men (i.e., more *sources* were identified as women as compared to men), men were quoted three times as often (i.e., overall *number of quotes* from males was far greater than the number of females quotes). Thus, men were given more space in each news article as compared to women. Women authors were more likely to quote other women as compared to males.	This methodology relied on predictable gender associations, which a) assumes gender is binary and b) are not always correct (i.e., 'Ryan' is not always male). Software sometimes accidentally pulled author byline information as part of the author's name, so manual data cleaning is required. Algorithms were trained on data that was written primary by men and quoting men; as a result, the software was more accurate in matching quotes to men as compared to women.
Jeon, 2020	Empirical	Perceptions of sexual sports violence	*Immediate purpose*: to provide information on the public's	1. A big data solution software (Textom) was used to pull text-based data from various social media websites.	The top 10 key words (ranking in terms of frequency) with 'sports sexual violence' were: 'sports circles',	Humans had to correct various errors in the data cleaning process (i.e., matching singular and plural versions

			perception of sports sexual violence. *Long-term purpose*: to draft improved policies on sports sexual violence and inform practical implications.	2. Authors searched for text using the key words 'sports sexual violence' 3. The top 50 key words associated with 'sports sexual violence' were identified. 4. Information was given based on how frequently each of the 50 key words were used, and the percentage of the key words' usage in relation to 'sports sexual violence'. 5. Errors were identified, and data was cleaned by humans. 6. Data was transformed into a matrix (a structure necessary for semantic network analysis). 7. Semantic network analysis was conducted to identify how the various key words were associated with 'sports sexual violence' and with one another. 8. Cluster analysis was conducted to assess how various key words hung together in groups. 9. Names were given to each group based on past research and the understanding of sports sexual violence.	'eradicate', 'measure', 'violence', 'National HR Commission', 'announce', 'investigate', 'victim', 'KSOC', and 'coach'. The key words that had the most influence were: 'sports circles', 'violence', 'eradicate', 'investigate', 'measures', and 'victims'. Those influential key words were then, in turn, most closely related to the key words: 'announce', 'Ministry of Culture', 'Sports and Tourism/MCST', 'Suk-hee Shim', and 'government'. Thus, a central theme related to sports violence was that the public wanted to see an investigation into the issues, punishment and response, and for legislation to be enacted. Eight groupings of key words were identified, which were named as: background of establishing a dedicated organization; government response; installment of national human rights commission; corruption in ice skating sports; government commitment; expanded investigation into the sports circles; accusation; and basic rights. Thus, the public believed it was important to address the causes of sports-related sexual violence, and that setting up responses to eradicate and prevent it was critical.	of words; correction for synonyms; removing duplicates). Data was limited to sampling from individuals who are accessing and using social media, and to those who do not face barriers to making their opinions on sports sexual violence known.
Moraes et al., 2011	Empirical	Physical intimate partner violence during pregnancy and the postpartum period	*Immediate purpose*: to identify the risk of mothers experiencing physical intimate partner violence during pregnancy and/or the postpartum period. *Long-term purpose*: to direct resources to mothers who are most at-risk.	1. Mothers with children under 5 months were interviewed to collect information on their pregnancy, post-partum period, children, partners, demographics, socio-economic status, lifestyle, and experience of physical intimate partner violence. 2. Authors categorized physical intimate partner violence (PIPV) as follows: no PIPV, PIPV during pregnancy OR post-partum period, PIPV during pregnancy AND post-partum period 3. Authors examined which factors were significantly associated with PIPV ($p < .05$). 4. Significant factors were included in a multinomial logit model.	Seven factors significantly predicted PIPV: perception of the child's health, maternal age, maternal schooling, number of offspring under 5 years old, mother's tobacco use, and alcohol misuse of the mother or the partner. These seven factors could be used to determine a spectrum of risk of experiencing PIPV during pregnancy and/or the post-partum period.	Authors were not able to include all relevant factors into the model; for example, there was no information collected on gestational age at the beginning of prenatal care. The decision of which factors to be included in the final model was driven by significance levels, and not a-priori hypotheses grounded in theory; this practice can lead to both type I and type II errors (reviewer original input).

				5. When factors lost significance in the full model, they were tested sequentially to determine which factors should be retained vs. removed. 6. A final multinomial logit model was identified, wherein all factors were statistically significant. 7. Ten scenarios were created to describe various situations of varying risk, based on the factors included in the multinomial logit model. 8. Risk for experiencing PIPV within each scenario was identified.		
Schneider, 2020	Commentary	Inequities in rental housing decisions	*Immediate purpose*: to discuss how algorithms can be used to prevent disparate treatment regarding fair housing. *Long-term purpose*: to ensure the field of data science follows best practices to promote social justice.	*Method for standard practice:* 1. Data is collected about an individual applicant (from proprietary records and social media companies). This data includes information such as eviction histories, criminal histories, prior addresses, internet browser histories, identity fraud, and credit data. 2. Machine learning is used to create an algorithm to predict the likelihood of one applicant paying rent on time or abiding by lease as compared to another. *Additional steps to ensure practice prevents disparate treatment:* Train the algorithms with data that reflects the society we want to live in (i.e., no disparities in criminal histories across race). Since these disparities exist in the real world, this step requires the creation of "fake data". Make all algorithms transparent and publicly available so others can review, critique, and suggest changes that need to be made to prevent disparate treatment.	If we can train algorithms with data that reflects an equitable world (i.e., no disparities across race, religion, gender etc.), and then use these algorithms to make housing decisions, we can promote social justice.	Often, algorithms are trained on data that reflect the true biases and disparities that exist in the world regarding race, religion, gender, etc. Thus, the algorithms reflect and perpetuate social inequities. Data on criminal histories and eviction histories are often incorrect. When the data is incorrect, the algorithm therefore makes decisions that are also incorrect.

4.0 Findings: How are data science methods being used to promote social change?

We intended to focus our review on studies that provided examples of how data science was being used to promote social change, particularly with regards to engaging men for violence prevention and the advancement of gender equality, diversity, justice, and inclusion. However, no studies explicitly engaged men in relation to these issues. Thus, the studies reviewed here broadly focused on using data science to promote social change. The studies focused on the following social issues: sexual violence[5,] physical intimate partner violence[6], gender equality[4], inequities and social justice in education[3], and inequities and social justice in housing decisions[7].

To help the reader fulsomely understand the complex data science methodologies, this review begins by providing an overview of *how* these methodologies are being used to promote social change. Specifically, we provide information on the various purposes the methodologies serve, along with the topics they focus on. We purposefully begin by providing this contextual information in the hopes that the detailed descriptions of the methodologies (i.e., actual steps taken to conduct the data science methods, described in 5.0) are more easily understood, as they are multifaceted and intricate.

4.1 Creating accurate characterizations of society

Two of the articles[4, 5] utilized data science methods for the purpose of accurately portraying what was happening in society. One study utilized big data to describe gender bias regarding how frequently women[iii] were quoted as compared to men in news reports[4]. Authors were able to examine this across several news sources, every day, for years. To this end, the purpose was to highlight gender bias as it occurs at any given time, but to also allow for the researchers to examine trends over time, and how gender bias might be impacted by historical events (i.e., COVID-19 pandemic). Although the immediate purpose of the methodology was to provide an accurate description of what is happening in society, the end goal was bigger than that. As stated by the authors, "By quantifying and measuring regular progress, we hope to motivate news organizations to provide a more diverse set of voices in their reporting" (p. 1). Thus, using big data to describe gender bias was a means to an end (changing society's behaviour).

The other article[5] utilized big data to describe the public's perception of sports sexual violence in response to the Korean Olympic gold medalist speed skater (Suk-hee Shim) speaking out about the sexual violence she experienced from her coach. Thus, the authors focused specifically on the societal response in Korea immediately following the date Shim publicly exposed her coach. To do

---

[iii] We use the terms "women" and "men" to reflect the focus of the original research, as discussed by authors. However, we wish to acknowledge that the authors attempted to also include those who identified as non-binary but were unable to due sample size and limitations of their software. We also wish to acknowledge that gender is non-binary, and that gender identities include both cis-gender transgender men and women.

this, they utilized big data from social media platforms, since "it enables interaction more than general media that delivers information, and may include perceptions such as reflection or critique by the public audience on the topic being delivered" (p. 29). Although the immediate purpose of the methodology was to provide a description of the public's perception, the ultimate goal was to draft improved policies on sports sexual violence and discuss practical implications of the research.

Big data is particularly well suited for the purpose of creating accurate characterizations of society, as it utilizes large data sets to answer questions from a social environmental perspective—as compared to research that focuses on individuals' perspectives.

4.2 Identifying inequity and/or promoting social justice

Three of the articles[3, 6, 7] utilized data science methods for the purpose of identifying inequities and/or promoting social justice. One article[3] described how data science methods can be used to promote social justice in education. Specifically, they discussed various ways in which big data and learning analytics could identify students who are struggling/disengaged/at-risk of poor academic outcomes (grades, dropping out), and then provide them with personalized learning that tailors to their unique needs. This is critical, as learning environments typically teach to the "average student", meaning that students who are disadvantaged in some way (e.g., students who face additional obstacles and barriers to learning) often "fall through the cracks". It can be difficult for teachers to provide personalized learning for each individual student, especially when class sizes are large (i.e., hundreds of students in a university course). By utilizing data science, the work of identifying students who are at-risk *and* providing them with personalized learning can be done automatically, greatly reducing the burden of teachers.

Although the immediate purpose of the methodology described was to identify students at-risk and then provide them with personalized learning, the end goal was to reduce social inequity. As the authors state, learning analytics-driven technologies have "potential to be a positive force for social change, and they are poised to have a central role in reduction of social inequity because they represent a unique, scalable learning technology that can be used to address *individual* students' learning needs" (p. 38).

Another article[6] utilized data science methods to understand what factors were most important to consider when predicting mothers' experience of physical intimate partner violence (PIPV) during pregnancy and the post-partum period, and then to also use those factors to predict risk across various scenarios. Although inequity was not explicitly discussed by the authors, we group this study here as authors utilized this methodology to identify which mothers were most at-risk for experiencing violence (thereby shining a light on inequities), and then discussed how resources could be directed to the mothers who are most at-risk (thereby moving one step toward reducing inequities). Therefore, although the immediate purpose of the methodology was to identify mothers at-risk for experiencing PIPV, the end goal was to provide those women with the specialized care they need to reduce their risk, and therefore reduce inequities.

The third article[7] discussed how data science can create algorithms that prevent disparate treatment with regard to fair housing. For context, the article discussed the "Fair Housing Act", which was created to ensure that people in the United States could not refuse to sell houses or rent property on the basis of race, color, religion, sex, or national origin. In an attempt to prevent disparate treatment, people have attempted to use data science to have computers make housing decisions, removing people (and therefore bias) from the equation. Specifically, algorithms were created to try and predict if one housing applicant is more likely than another housing applicant to pay rent and/or to abide by the lease. The authors critically analyzed this methodology and highlighted how algorithms can indeed be created to do this and how they can remove bias from the equation. Perhaps more importantly, however, they also discussed how dangerous this methodology can be –and that most often, creating algorithms based on real data (the current practice) often reflects the real biases and racism that exists in the world. By highlighting the importance of critically thinking through data science methodologies (discussed in detail below, see 5.2; 6.1.1; 6.1.2), the authors discussed how algorithms can be used correctly to promote social justice, but also how they can be used incorrectly to perpetuate social inequities.

Big data is particularly well suited for the purpose of identifying inequity and/or promoting social justice, as it utilizes large data sets that are often more diverse, and thus more closely representative of the true population—as compared to research that focuses on smaller sample sizes that are often impacted by sampling bias and convenience sampling, resulting in homogenous samples that are not representative of the larger population.

5.0 Findings: What are the key data science methods?

All of the articles utilized different methods in data science. For purposes of this review, we group the methods according to similarities in process and/or purpose. Specifically, we discuss data science methods that a) collect and analyze text-based data from publicly available online sources to extract meaning, and b) use predictive modeling to predict risk and/or behaviour. It is important to note that a plethora of data science methods exist, and the articles reviewed here provide only a very small sampling of the methods available.

Given the complexity of the methodologies detailed in the empirical articles[4, 5, 6], which involved multiple steps, we interweave the results of the empirical studies within this section. We do this purposefully to help the reader fulsomely understand the method itself, and how it can be used as a tool to promote social change. The empirical articles provided more details as compared to the commentaries, which only gave general overviews of the methods and what they can do. Thus, differences in details of the methodologies and the inclusion of results will differ depending on the type of article.

5.1 Text-based methods

Two of the studies[4, 5] utilized methods that collect and analyze text-based data from publicly available online sources. To this end, the methods included the following (very simplified) steps:

1. Using software to pull text from various sources (news articles, social media).
2. Cleaning the data to ensure it is meaningful for the purpose of the study and to check for errors.
3. Analyzing the data to identify trends that give meaning to the text.

However, each method completed this process in different ways, as described below.

The purpose of one study[4] (discussed above, see 4.1) was to compare how many times women were quoted in news stories online as compared to men. To do this, the authors used *large-scale text processing* and *big data storage* to pull the text from various news sources. The authors used a software called "Gender Gap Tracker", which they created specifically for this study. Since every news source has a unique layout for their website, the method of pulling text had to be tailored to the individual website. Thus, the code for the software had to be written differently for each unique website, which was an arduous process. Once the code was written to pull the text-based data, the software pulled the data from the websites twice a day, and then transferred the data to a central database. Each process of pulling the data took between 5 and 30 minutes to complete.

Once the full text data was imported into their central database, the authors completed additional steps to pull quotes from the full text data. Specifically, they pulled both direct quotes (somebody's exact words, as indicated by quotation marks, for example *"I am using data science"*) and indirect quotes (when someone's words are altered in some way, for example: *she was using data science*). The software then identified the source of the quotes, and who each quote belongs to.

Following this, the software used natural language processing to a) identify the sources of the quotes as well as the authors of each article, and b) predict their gender. To do this, they utilized Named Entity Recognition (NER), which is a machine learning process for classifying text into categories (here, the categories are: male, female, other). Since this process relies on "predictable gender associations" based on people's first and last names (i.e., Robert is typically male; Amelia is typically female), and because it assumes gender is binary, there are inherent limitations of this method that require careful evaluation and data cleaning.

Often, it is necessary to link the source's names to other available information (i.e., Wikidata) to confirm someone's gender identity. However, this is not always possible to do, so there are likely errors in the data. Furthermore, sometimes the software pulls an article's author information from the byline (such as 'Posted', 'Last Updated') in addition to their names, which requires manual data cleaning to correct. In addition, it's necessary to consistently evaluate this process to make sure errors are not being made to disproportionately attribute names or quotes to one gender (discussed in detail in section 6.1.1).

Once all quotes are identified, all sources are identified, and all sources are assigned to a gender, the analyses can begin. For purposes of their project, authors analyzed counts and percentages of female vs. male sources across seven news outlets (CBC News, CTV News, Global News, HuffPost Canada, National Post, The Globe and Mail, The Star), and then counts and percentages overall (all sources combined/aggregated data). As mentioned previously, this data was collected daily. The information regarding counts and percentages are updated every 24 hours and presented on a data dashboard. This way, the public can interact with the dashboard to see the results for any day (starting October 1, 2018). The authors found that more women were used for quotes as compared to men (i.e., the number of women identified as *sources* was greater than the number of men identified as *sources*). However, when the authors examined the number of male *quotes* vs. the number of female *quotes* in the articles, gender bias emerged. Specifically, even though there were more sources identified as women, *the actual number of quotes by men outnumbered the number of quotes by women*. In fact, men were quoted three times as often as females. Thus, the authors concluded that men are given more space than women in online news media. Additionally, female authors were more likely to quote women as compared to men. As such, the authors suggested part of the solution to reduce the gender gap in quotes is to have more women reporters.

In sum, this methodology allows people to use a software system to collect text-based data from the internet, extract quotes, assign gender to the quotes, and analyze the frequency of quotes for women as compared to men.

The other article[5] that used text-based data science methodology aimed to assess the public's social perception of policies on sports sexual violence. To do this, the authors used Textom (a big data solution software) to pull text from various social media websites. Then, the authors selected a set of key words to search for. For purposes of their study, the authors searched for text using the key words 'sports sexual violence' on a select set of social media websites during a specific time frame. From that initial data set, one can pull any number of words that are mentioned in relation to 'sports sexual violence'. Given the copious amount of data available, and the difficulty in visualizing and interpreting big data, the authors limited their search of associated key words to 50. Thus, the authors were able to pull the top 50 key words that were mentioned along with 'sports sexual violence'. To this end, the data contains information on what the key words were (top three key words were: 'sports circles', 'eradicate', and 'measure'), the key word's ranking (ranking from 1 – 50, 1 = most frequently used key word with sports sexual violence), how many times the key words were used in relation to 'sports sexual violence' (frequency, numeric value), and percentage (out of all words used in relation to 'sports sexual violence', what percentage of the time is that key word used?).

Once the data was extracted, it had to be cleaned. Authors indicated that various errors were corrected (i.e., matching singular and plural versions of words; correcting for synonyms) and that duplicates needed to be identified and removed. Then, data is transformed into a matrix (which is a way to structure the data within rows and columns that show each variables' relation to one another in a meaningful way) for analysis.

Once the data was cleaned, the analyses took place. Authors first conducted a *semantic network analysis*, which identifies how the various key words were associated with the term 'sports sexual violence' and with one another. This was done by identifying which key words played the most central role, and which key words had the most influence (serving as an agent among other key words). The results of this analysis are often displayed as a visualization. To aid in the understanding of this visualization, see Figure 3 below, taken directly from the article on sports violence[5] (https://www.ajol.info/index.php/ejhd/article/view/198772). Each key word is represented with a square, and the relationship between each key word/square is represented with a line. The larger the square, the more central that key word is to 'sports sexual violence'. The more lines going to and from the square, the more influential the word is. The words that had the most impact were 'sports circles', 'violence', 'eradicate', 'investigate', 'measures', and 'victims'. These words were closely related to the following list of words: 'announce', 'Ministry of Culture', 'Sports and Tourism/MCST', 'Suk-hee Shim', and 'government'. Thus, the authors concluded that a central theme related to sports sexual violence is "the investigation into the issues, the establishment of punishment and response, and the enactment of legislation on this issue" (p. 35).

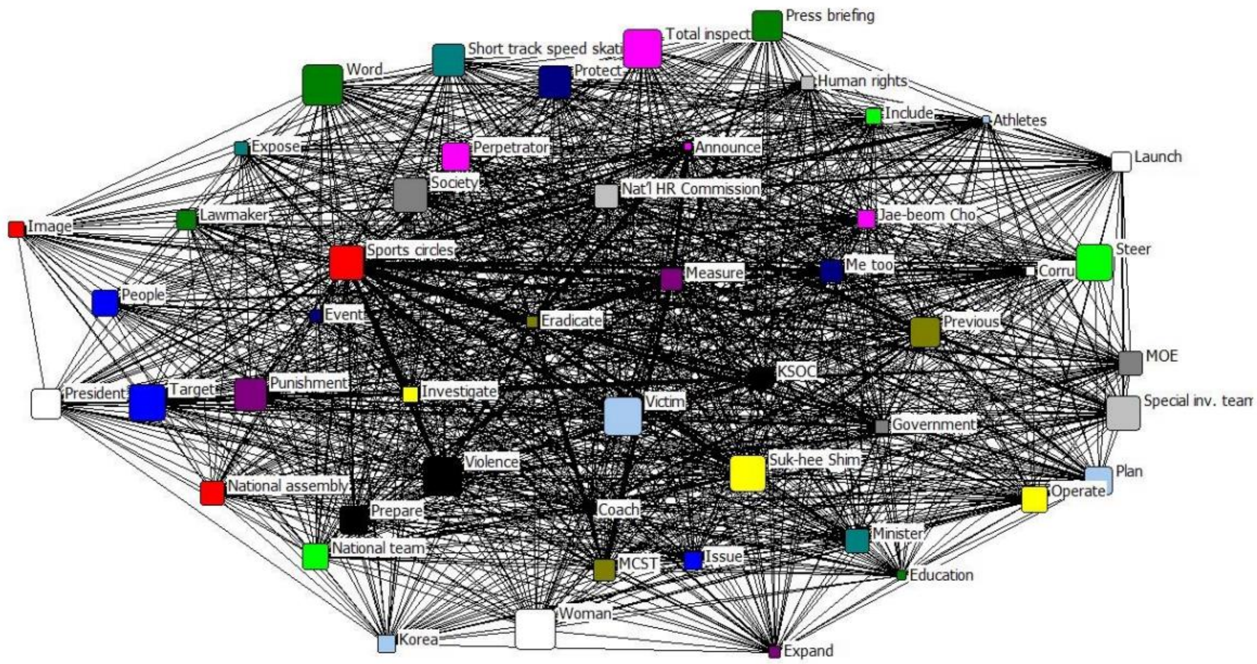


Figure 3. Example of semantic network analysis visualization

Second, the authors conducted a cluster analysis, which is a way to assess which key words "hang together" in groupings. To help in the understanding of this visualization, see Figure 4 below, taken directly from the article on sports violence[5] (https://www.ajol.info/index.php/ejhd/article/view/198772). Each cluster was then given a name (by the researcher) based on the understanding of past research and knowledge related to the study. As shown in Figure 4, eight groups were identified and named by the authors as follows: background of establishing a dedicated organization; government response; installment of national human rights commission; corruption in ice skating sports; government commitment; expanded investigation into the sports circles; accusation; and basic rights. Here, we highlight two examples: Group 7 below contained the key words 'expose' and 'me too' and was thus named "accusation"; group 8 below contained the words 'human rights', 'protect', and 'violence', and was thus named "basic rights".

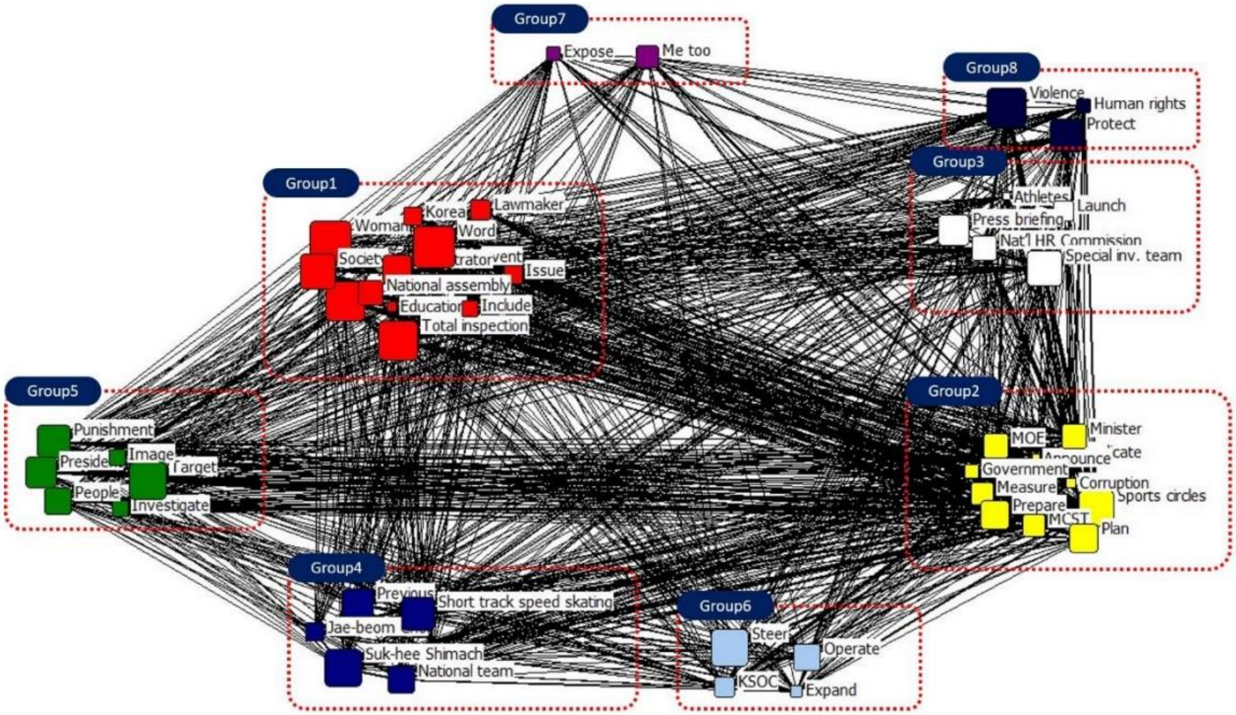


Figure 4. Example of cluster analysis visualization

In sum, this methodology allows one to use a software system to collect text-based data from social media, extract specific key words relevant to the subject of interest (i.e., sports sexual violence), extract other key words that are associated with the subject of interest, and analyze patterns for how people are "talking" about it. From these analyses, authors concluded that the public believed it was important to address the causes of sports-related sexual violence, and that setting up responses to eradicate and prevent sports sexual violence was critical.

5.2 Predictive modeling

The other three studies use some form of predictive modeling to predict people's risk or behaviour. Although the unit of analysis for predictive modeling can take many forms, these studies analyzed data at the level of the individual (as compared to the text-based models that focused on the social environmental level).

One study[6] used data science to understand what factors could predict[iv] women's risk of experiencing physical intimate partner violence (PIPV) during pregnancy and/or the post-partum period. The authors randomly selected mothers who had children under the age of 5 months that were in primary health care waiting rooms in Brazil. 811 mothers were interviewed about a) their experience of PIPV during pregnancy and/or the post-partum period, b) characteristics of their

[iv] We use the word "predict" to be consistent with the author's terminology. However, we would caution the reader as their methodology used cross-sectional, retrospective interviews, which cannot infer causation.

children (e.g., child's age, sex, gestational age), c) their [mothers'] characteristics (e.g., age, education, race), d) their [mothers' and partners'] lifestyle (e.g., tobacco and alcohol use), and e) socio-economic status (e.g., amount of household goods, occupation of family's main income earner). The authors categorized PIPV experience as follows: no PIPV, PIPV during pregnancy OR post-partum, PIPV during pregnancy AND postpartum. Multinomial logit models (a type of regression analysis that predicts categorical outcomes) were used to predict mother's likelihood of falling into one of these three categories. To do this, authors first examined whether each factor (i.e., child's age, mother's alcohol use) was significantly associated with PIPV ($p < .05$). All factors that were statistically significant were considered in the multinomial logit model. Then, the factors that were statistically significant were entered into the multinomial logit model. Since some factors lost significance when entered into the multinomial model (i.e., factors that were significantly associated with PIPV by themselves, but not significant when considered along with other factors), the authors tested those factors sequentially to determine which factors should be removed. The goal was to have all factors in the final model be statistically significant. Thus, the technique here was led by sequential testing (by humans) and no artificial intelligence/machine learning algorithm was used.[v]

The authors found that seven factors significantly predicted PIPV: perception of the child's health, maternal age, maternal schooling, number of offspring under 5 years old, mother's tobacco use, and alcohol misuse of the mother or the partner. Once a final model was determined, the authors then created ten "scenarios" of what a mother's experience could look like based on different combinations of these seven factors. The scenarios were meant to portray various levels of risk. For example, the first scenario is the "lowest risk" scenario and consists of a mother who has excellent perceptions of her child's health, is over 20 years old, has a medium level of maternal schooling (or greater), only one child, does not use tobacco, does not use alcohol, and does not use drugs. Compare that to the tenth (last) scenario, which is the "highest risk" scenario which consist of a mother who perceives her child's health to be less than excellent, is under 20 years old, has not completed a medium level of schooling, has two or more children under 5, smokes, uses alcohol, and drugs. The authors then showed how mothers falling into these various scenarios differed in terms of risk of PIPV. As expected, the lowest risk scenarios had the lowest risk of PIPV, and risk of experiencing PIPV increased as the scenarios increased in risk.

In sum, this methodology allowed the authors to determine which factors were most important to consider when screening women for risk of PIPV, and to describe how risk could increase across varying scenarios.

Another article[3] described how learning analytics can use large amounts of data to predict which students are at-risk for poor academic outcomes, and then use that information to provide them with personalized learning to address their unique needs. Although this article was a commentary, and thus did not provide detailed information on methodology, general information was given

---

[v] Please see the limitations section for a discussion of how this practice of letting the data decide (without a-priori hypotheses) can be dangerous.

regarding how this is done. First, data is collected from the students. Often, administrative data is used that gives information such as student demographics, student grades, and students' behaviours (i.e., class attendance). Importantly, this data is collected from all students across a certain learning environment. Thus, this could be done across an entire campus, for example, resulting in a very large amount of data. Then the data is analyzed to predict which students are most at-risk for low grades, dropping out, etc. Importantly, with such a large amount of data, it is also feasible to look not only at individual students, but how trends might differ within and across students, courses, or populations. This information is often displayed on data dashboards, which are visuals that show snapshots of information about the students in real-time. These dashboards can be used by faculty and staff (teachers, professors, guidance counselors) to identify students at risk; they can also be used by the students themselves so they can monitor their own standing and progress. Importantly, learning analytics-driven technology can then offer suggestions for how to tailor students' learning. In sum, this methodology uses big data and learning analytics to identify which students are most at-risk of poor academic outcomes, and then provide them with learning that is personalized to their needs at that time.

The third article[7] described how algorithms can be used to make automated housing decisions. The authors discuss how various companies exist that use algorithms to assess if one applicant for an apartment is more likely than another to pay rent on time or abide by the lease. To do this, data is collected about the individual applicant. Often, the following data is collected (from proprietary records and social media companies): eviction histories, criminal histories, prior addresses, internet browser histories, identity fraud, and credit data. Then machine learning models create an algorithm to predict the likelihood of someone paying rent on time or abiding by the lease. Importantly, the authors highlight the fact that these algorithms are a "black box", and it is not known how the decision is made. The authors subsequently argue that when this is the case, the algorithms can't be trusted. The authors discuss a variety of limitations of this methodology (detailed below, see sections 6.1.1 and 6.1.2), and then propose suggestions for how algorithms can be created that are trusted. Mainly, the information used to create the algorithms should be transparent, so others can evaluate the ethics behind it. Second, the algorithms should be trained with data that reflects the society that we want to live in. This means that the algorithms should be trained with data sets that do not show any disparities in the variables (eviction histories, criminal histories, etc.) across race, gender, religion, etc. If algorithms can be trained on truly unbiased data, then the algorithms themselves will also be unbiased. This data is, unfortunately, non-existent since we live in a world rampant with disparities. Thus, it is imperative for people to build these "fake" data sets for the algorithms to learn on. In sum, this article described how data science can be used to make fair housing decisions that would ultimately promote social justice.

6.0 Findings: What are the key considerations and cautions?

Thus far, we have provided an overview of how data science is being used to promote social change, and the details behind the methodologies. Little critique has been given to the key considerations and cautions when utilizing data science for this purpose. Here, we review key themes that emerged during the review, to ensure the reader is aware of limitations of these methodologies. Some themes were explicitly stated in the articles. However, additional considerations and cautions were provided by the first author. We feel it is important to note that all methods, but particularly data science methods, need to be fully understood before they are used. Often, researchers know "just enough to be dangerous", or that researchers often know the basics of how a methodology works but may not have a fulsome understanding that enables them to think critically about the potential limitations of their own research. For this reason, we strongly recommend researchers thoroughly educate themselves on each methodology before use and/or collaborate with experts in the field to ensure these methods are used ethically, and that results are presented in the context of these limitations.

6.1 Data fundamentalism

Broadly, we first discuss the issue of *data fundamentalism*, which refers to the "notion that correlation always indicates causation, and that massive data sets and predictive analytics always reflect objective truth[8]". This notion is false, and it is imperative that anyone working with data has a fulsome understanding of how data is *not*, in fact, objective. In addition to the recommendations discussed here, we strongly recommend people in the field of engaging men educate themselves on the Data Equity Framework created by We All Count[9]. The framework helps people understand how each and every decision that is made with data (from sampling to analyzing to interpreting) is influenced by people's biases and can therefore bias results. Importantly, We All Count provides a set of tools, checklists, and practices that can be used to help make decisions intentionally to move towards data equity.

Here, we divide the discussion of data fundamentalism into two separate (but related) components: the data sets themselves, and then the analyses used to create the end results.

6.1.1 Data sets

For data sets to reflect objective truth, they must be accurately representative of the world we live in, and they must be without error. This is never the case. Indeed, three[4, 5, 7] studies mentioned the importance of identifying errors and manually cleaning the data. For example, when methods are used to scrape text-based data from the internet, errors are often made. As was the case when scraping the text from news stories to compare women vs. male representation, the software sometimes accidentally pulled an article's author information from the byline (including text like 'Posted', 'Last Updated') in addition to their names[4]. In doing so, the data set contained incorrect information regarding the authors' names –and therefore the authors' genders. Thus, manual data cleaning was required to ensure the data collected is truly reflective of the author information. Without this attention to detail, the data set is inherently "wrong".

Relatedly, software is not able to assign gender correctly 100% of the time. That is because as of now, software assumes that gender is binary, that sex and gender are perfectly correlated, and that names can be used to accurately predict gender -again, this is not the case[4]. Thus, the data sets created are often limited to only include data that is more easily understood/not prone to as much error (males vs. female; non-binary identities excluded), and even then, mistakes are made (i.e., 'Ryan' may be classified as male, when in fact the author is female). Therefore, in this instance the data is again “wrong”. These methodologies require humans to oversee the process by which data is collected and cleaned, as software is not able to pay attention to these nuances and make decisions accordingly.

As another example for pulling and cleaning text-based data, software is not always able to make decisions regarding singular/plural versions of words, abbreviations, synonyms, and antonyms[5]. Again, humans are required to oversee this process to ensure the data is accurate. However, even when humans oversee the process, it is impossible to guarantee the data sets are completely accurate, given their massive size.

Another article discussed this explicitly, highlighting the fact that data that is pulled on criminal histories and evictions are often inaccurate[7]. Thus, if decisions are made based on inaccurate data, then the results will again be “wrong”.

Furthermore, in an ideal world, data sets are samples that are representative of the larger population. However, this is a peril of all research involving human subjects. We can never ensure we have data from everyone in the population, or that our subset is truly representative of everyone in terms of demographics, geographic location, childhood experiences, etc.

For example, when pulling text-based data from social media sites, the data is only representative of individuals who a) use social media and b) actively speak up about the topic the researchers are interested in. Consider the example of using social media to analyze perceptions around sports sexual violence[5]. This methodology would not pull data from individuals who have issues with accessibility and awareness of social media, which may result in a data set that is ageist. Furthermore, there are many reasons why individuals may choose to not speak up about this issue. For example, some women may fear a negative response to their posts due to culture of victim blaming; some men may not want to express empathy due to harmful gender norms that perpetuate the idea men who show empathy are weak. Thus, the data collected from social media sites only sample from a subset of the population that has strong opinions on sports sexual violence and also does not face barriers to speaking up about it.

In other scenarios, data is pulled from a relatively small sample, as was the case when examining mother’s risk of experiencing PIPV[6]. Other than the fact that the sample was limited in size and limited to one geographical location, it is also possible that mothers who were experiencing severe abuse were unlikely to participate in the study, perhaps due to fear of partners finding out about their participation.

Finally, it is important to note that no matter how comprehensive the data set, it will never include “everything”. Indeed, two studies explicitly mentioned the fact that it is impossible to measure everything[3, 6]. As one example, when predicting risk of PIPV, the data set did not include important factors like gestational age at the beginning of prenatal care, of the age difference between the mother and her partner[6]. As another example, administrative education data often doesn’t have information regarding dimensions of student learning, which are important when assessing risk for poor academic outcomes[3].

In sum, it is important to note that the actual data set will never be "correct", in that it will not be free from error, it will not be completely representative of the population of interest, and it will not contain information on every variable that is relevant to the topic of interest.

6.1.2 Analyses

Analyses in data science are often conducted with very large data sets (big data). Although this is advantageous in many ways, it also presents a problem. It is very easy to "find" results that are statistically significant when the sample size is very large; however, these significant results are not always practically significant or meaningful[10]. Said another way, it's easier for analyses to result in Type I errors (when you find a result that is significant according to the analysis [i.e., p value < .05], but the result is not "truly" significant in the real world). This is an artifact of the null hypothesis testing procedure (for a detailed explanation, see Khalilzadeh & Tasci, 2017). Since researchers are likely to find statistically significant results "by accident" when using big data, it is critical for researchers to be aware of this and to interpret their findings in context. One way to do this is by estimating an effect size, which tells you how strong an effect is[10]. With this information, one can draw conclusions about whether the results of the analyses are also practically significant (e.g., does the statistically significant finding "matter" in the real world?).

Thus, it is essential for researchers to think critically about the analyses they plan to run, and the hypotheses they have for each analysis. Indeed, researchers have argued that analyses conducted with big data needs to be guided by theory[11]. Although it can be tempting to use data science methods to identify trends in the data, this is a dangerous practice if it is conducted without the guidance of theory and/or a thorough understanding of the research in that area. Correlation between two variables does not imply causation. Thus, when a finding is detected, researchers need to think critically about whether it was expected or not. As discussed previously, big data is prone to finding spurious results. Furthermore, it is impossible to collect data on every single variable of interest. Thus, a correlation between two variables may in fact be driven by a third (unmeasured) variable.

For example, consider a hypothetical scenario in which researchers are investigating a large data set that contains information on reports of suspected child abuse during the COVID-19 pandemic. If running analyses "blindly" (without theory or a thorough understanding of the research in this area), a researcher may find a statistically significant drop in how often referrals to child protective services were made in October 2020. The researchers may conclude that incidents of child abuse dropped during that time, perhaps due to improved family life with everyone at home together due to stay-at-home mandates. However, researchers in this area know the drop in referrals is actually reflective of the fact that children were not in schools during this time, which is where referrals are commonly made by school staff. Without that understanding, the researchers might have presented information and interpreted it incorrectly. Thus, the use of theory and understanding of previous research and context is critical, highlighting the importance of having a-priori hypotheses.

For this reason, there is significant caution in researchers letting the data drive the decisions for what variables should be considered important (often, decided to be important if the variable is statistically significant in the model). When researchers try to find results that are statistically significant, they will be successful. However, they will not always be correct in their interpretations. This is also dangerous, as it impacts interpretations. As an example, if a data-driven model decided that tobacco use was a predictive factor for experiencing violence, a public health specialist may conclude that tobacco prevention is needed for violence prevention. However, what is likely needed is an investigation into the root causes for such coping strategies, which would lead to a more informative avenue for violence prevention.

Some of the reviewed articles discussed ways to attempt to mitigate the risk. For example, the study that examined sports sexual violence[5] used a peer consultation method where the researcher conducted iterative consultations with researchers in a variety of areas (public health management, sports sociology, big data) to result in a more objective interpretation of the results. However, an explicit discussion of how theory was used to guide the research questions, hypotheses, analyses, and interpretation was missing from the reviewed articles. Thus, we reiterate the importance of building oversight into each research project.

One final note, the software/analyses/algorithms that are used to analyze the data will be biased if the data they (software/analyses/algorithms) are trained on is biased. This was mentioned in two articles[4, 7]. The first used software to determine which quotes belonged to women and which quotes belonged to men[4]. The software was trained on data that was written *primarily by men and quoting men* (and therefore biased towards men). Thus, it was unsurprising that the authors found the method's ability to match quotes to men was more accurate as compared to the accuracy of the method's ability to match quotes to women.

The second described algorithms that are used to predict if one individual is more likely to pay rent or abide by the lease as compared to another individual[7]. As mentioned previously, this algorithm is based on data such as eviction histories, criminal histories, and prior addresses. However, data sets with this information reflect the biases that exist in the real world. For example, since the criminal justice system reflects disparate treatment of minorities, a data set on criminal histories will be biased against minorities. If algorithms are trained on data that contain the biases that are reflected in the real world, then the algorithms (and therefore findings/implications) of this research will perpetuate those biases and continue the cycle of disparate treatment. This is extremely dangerous practice that has serious (and often irreversible) impacts.

We wish to repeat the importance of using a data equity framework, which can be used to "demonstrate that the data you are producing is reliable and most importantly created in a process that brings awareness, intentionality, and meaningfully embeds a strong definition of equity at each step"[9].

6.1.3 Privacy

Finally, two studies[3, 7] mentioned concerns of privacy when utilizing big data. First, in the case of learning analytics[3], students do not always consent to having their information used. Some learning dashboards automatically pull their information (test scores, class attendance, etc.) and present it in dashboards. This could, understandably, cause a student to have concerns of a lack of privacy. Furthermore, when learning analytics are used continually (throughout the entire course), it is debated whether consent given once (at the beginning of a course) can truly hold throughout the duration of the semester (or beyond). Second, another article[7] discussed the fact that although it is recommended that companies be transparent about the algorithms they use to make decisions, this is not always legally possible. Some organizations may have valid trade secrets or privacy concerns, making it impossible to be fully transparent about the data used to create algorithms. These issues are still being critically discussed, so no clear recommendations are provided here to address these concerns. Additionally, there have been instances where influential donors have requested information from the organizations they financially support, which would result in a breach of confidentiality and trust; thus, it's important to have procedures in place so that the safety and dignity of clients are set as the priority.[12]

7.0 Recommendations: What are key considerations and cautions for those in the field of engaging men?

The studies reviewed did not focus on the field of engaging men. Thus, we were unable to answer this question from the studies themselves. However, we can make concrete recommendations on how policy makers, funders, practitioners, and researchers in the field of engaging men might utilize these methods to advance the field. Thus, the recommendations here are reviewer original input.

7.0.1 Awareness and education

First, we recommend that those in the field of engaging men become aware of the field of data science and begin educating themselves on what it entails. Reading this review could be a first step in that process, but it will also involve a much broader education of methodologies used in data science. As mentioned previously, the articles reviewed here only represent a very small sample of all the methods available for use. This could be done by reviewing grey literature on data science, completing courses in data science, and/or collaborating with those in the field of data science.

7.0.2 Consider how data science can be used as a tool to mobilize men

Once there is an understanding of what data science is, and the various techniques used, one can consider the ways in which it could be applied to the field of engaging men. To this end, we hope people can recognize data science as a tool that can be used to move the field forward. Here, we illustrate one example of how this might be done.

As one (oversimplified) example, we could utilize text-based data science methods to answer the following question: *"Can we support fathers in Calgary to be positively engaged with their children?".* To do this, we could first scrape data from male-identified users on Twitter (in Calgary) that use some combination of the following key words: 'fatherhood', 'dad', and 'parenting'. We could then use the semantic network analysis described in this review[5] to analyze the themes related to how males are talking about fatherhood. Second, we could implement some sort of social media campaign to nudge fathers to become more engaged with their children. Following the intervention, we could repeat the process (collect data from Twitter, analyze for themes using semantic network analysis) to examine if there was any change in how men are talking about fatherhood. Based on the findings, we could then make modifications to the social media campaign to continually improve our approach. For example, if our analysis revealed a key set of positive words being used to amplify the value and importance of fatherhood, we could strategically use those exact words in the next social media campaign.

7.0.3 Increase and improve data collection

In order to utilize data science techniques, big data is needed. As of present, the data available to the field of engaging and mobilizing men is limited. Although it is possible to utilize data from the internet (a massive source), it would be ideal to have other large data sets with more detailed information, and information that is specific to the agenda of engaging men. However, most organizations either do not collect data, or there is room for improvement in their data collection. For example, there is no consistent method of collecting data across non-profit organizations that serve men or male-dominated workplaces (i.e., police agencies). Large amounts of data, and importantly, good quality data, are needed as a first step. Thus, we recommend funding and support surrounding the collection, storage, and usage of data within leaders and organizations that are trying to advance the field. We recognize this process will take time, as some individuals (and organizations) have mistrust regarding collecting and using sensitive data. It will require resources, education, trusting partnerships, new competencies, and time for this to happen. Promisingly, various efforts are being made to do this, and progress has already been made. For example, data is being consistently collected from Nova Scotia's Men Helpline (211), a helpline that engages men who are at-risk for using violence to assess, monitor, manage, and reduce abusive behaviours[13]. The data is then analyzed with the goal of informing prevention efforts with men. As another example, an Alberta initiative (co-led by Shift, HelpSeeker, and IMPACT) called *Data2Action*[14] is currently working to help the domestic and sexual violence sector learn how to systematically collect data on domestic and sexual violence across the province, and then use that data to create policy, systems and practice change to advance violence prevention.

7.0.4 Create safety nets

As a final note, the authors urge anyone engaging in data science to be aware of the limitations of the practice. Without understanding the limitations, data science may be used incorrectly, perpetuate biases, present false information, and cause irreversible harm. Given the in-depth knowledge that is needed to use data science methods correctly and ethically, we recommend (as a first step) that those in the field of engaging men partner with others who have this in-depth knowledge and experience. By doing this, a safety net is created whereby a diverse set of individuals can act as checks and balances to ensure that the methods are being used appropriately. We All Count[9] (mentioned previously, see 6.1) offers a variety of tools and services to help individuals and entire organizations to do this.

8.0 Conclusion

In sum, data science focuses on creating, collecting, handling, and analyzing large amounts of data with the purpose of extracting actionable insights. Data scientists use these methods to identify patterns and trends, predict future events, reduce risk, and improve programming and outcomes. Data science is an emerging practice for those interested in using it as a tool to promote social change. Thus far, it has been used to analyze trends in the public's perception of sexual violence; to highlight gender bias in news articles; to predict risk of physical intimate partner violence among mothers during pregnancy and the postpartum period; to prevent disparate treatment regarding housing decisions; and to identify students at-risk for adverse learning outcomes, then provide them with personalized learning. Data science is not yet being utilized in the field of engaging men. With careful consideration, we believe data science can be used to significantly progress and advance the field of engaging men. Although this will take both effort and time, we believe the long-term payoffs will be worth it.

THE PROJECT TO END DOMESTIC VIOLENCE

References

[1] *Data Equity Framework*. (2021, August 3). We All Count. https://weallcount.com/the-data-process/.
[2] Lee, L. & Wells, L. (2021). *Building Alberta's primary prevention framework: Primary prevention concepts* [PowerPoint Presentation]. Calgary, AB. The University of Calgary, Shift: The Project to End Domestic Violence.
[3] Dobbins, M. (2018). Rapid Review Guidebook: Steps for Conducting a Rapid Review. National Collaborating Centre for Methods and Tools. McMaster University.
[4] Aguilar, S. J. (2018). Learning Analytics: At the Nexus of Big Data, Digital Innovation, and Social Justice in Education. *TechTrends*, *62*(1), 37–45. https://doi.org/10.1007/s11528-017-0226-9
[5] Asr, F. T., Mazraeh, M., Lopes, A., Gautam, V., Gonzales, J., Rao, P., & Taboada, M. (2021). The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media. *PLoS ONE*, *16*(1), 1–28. https://doi.org/10.1371/journal.pone.0245533
[6] Jeon, S. (2020). A study on the social phenomenon and perception of sports sexual violence through big data analysis of social media. *Ethiopian Journal of Health Development*, *34*(3), Article 3. https://www.ajol.info/index.php/ejhd/article/view/198772
[7] Moraes, C. L., Silva, T. de S. T. da, Reichenheim, M. E., Azevedo, G. L., Oliveira, A. S. D., & Braga, J. U. (2011). Physical violence between intimate partners during pregnancy and postpartum: A prediction model for use in primary health care facilities. *Paediatric and Perinatal Epidemiology*, *25*(5), 478–486. https://doi.org/10.1111/j.1365-3016.2011.01208.x
[8] Schneider, V. (2020). Locked out by big data: How big data, Algorithms and machine learning may undermine housing justice. *Columbia Human Rights Law Review*, *52*(1), 55.
[8] Crawford, K. (2013). The hidden biases in big data. *Harvard business review*, *1*(4).
[9] *Data Equity Framework*. We All Count. (2021, August 3). Retrieved October 5, 2021, from https://weallcount.com/the-data-process/.
[10] Khalilzadeh, J., & Tasci, A. D. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, *62*, 89-96.
[11] Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2080), 20160153.
[12] Behnam, N., & Crabtree, K. (2019). Big data, little ethics: confidentiality and consent. *Forced Migration Review,* (61), 4-6.
[13] *Nova Scotia Men's Helpline Evaluation Report.* Retrieved October 5, 2021, from https://www.learningtoendabuse.ca/our-work/our-projects-resources/nova_scotia_mens_helpline/MHL-Final-Report-August-2021.pdf.
[14] Turner, A., Wells, L., McManus, C., & Baker, E. (2021). *Data2Action: Using Data to End Domestic and Sexual Violence* [Webinar]. Helpseeker. https://www.helpseeker.co/webinar/data2action/