Introduction

"More Gender Equality, the Fewer Women in STEM (Khazan, 2018)." Author Olga Khazan used the statement as a title for her piece at The Atlantic explaining the gender gap in STEM fields globally. The article cited The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education by Gijsbert Stoet and David Geary and used Figure 1 to support the argument made earlier in the title.

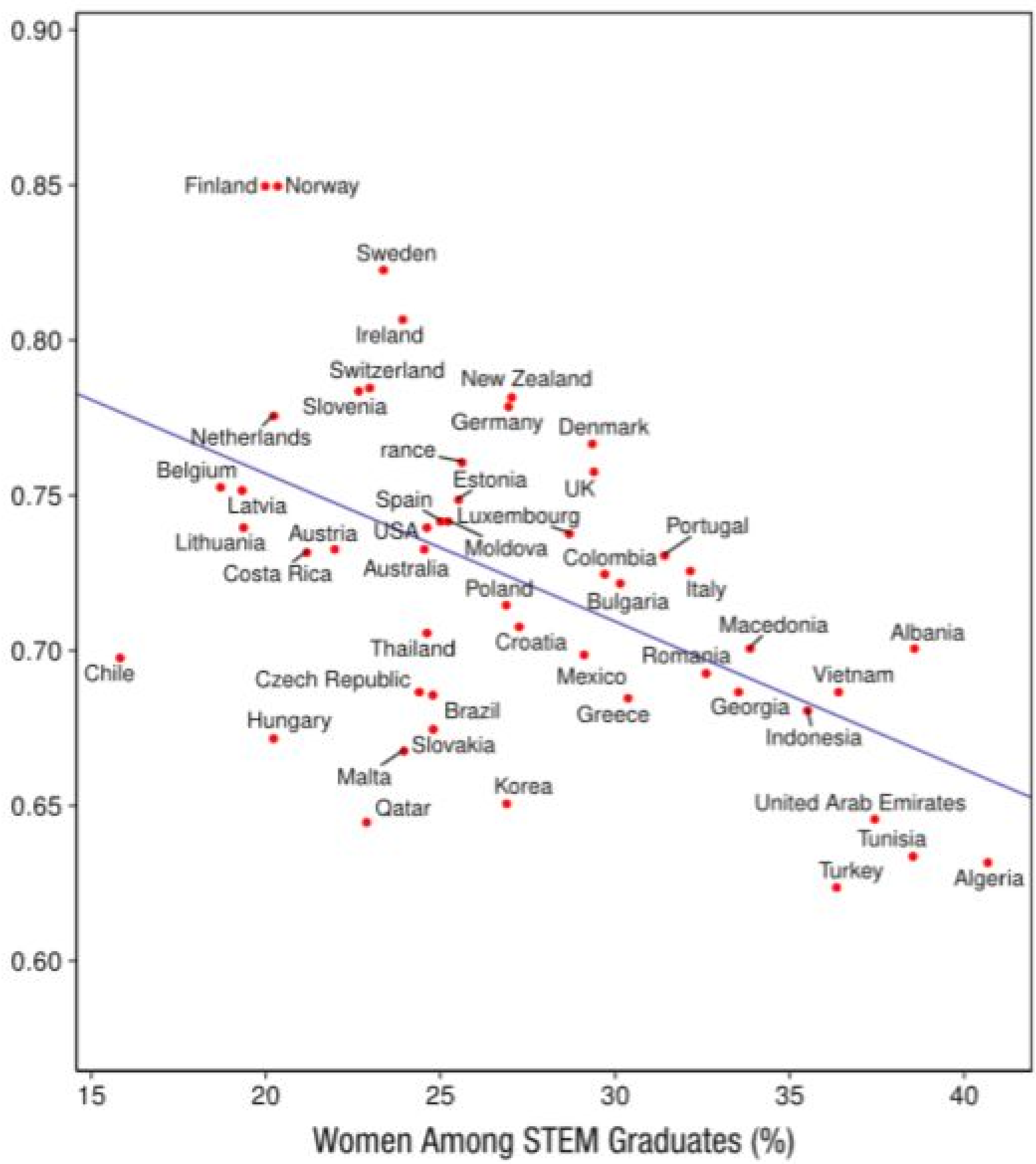

Figure 1. The relationship between Women Among STEM graduates and the Global Gender Gap Index (GGGI[1]) for 77 countries.

The authors in the paper used Figure 1 to conclude that being in a welfare state, which is correlated with a higher value of GGGI, motivates women to pursue non-STEM fields. This is a classic example of "correlation does not mean causation" and how flawed causal analysis can further solidify cultural and patriarchal norms about women's capacities in STEM fields.

In this piece, I explain how Stoet and Geary's approach towards causality is empirically flawed. Also, I explain how we can study the relationship between gender equality and the participation of women in STEM fields using Pearl's causality framework.

Pearl's Framework for Causality

Since correlation does not mean causation, it is important to understand what causation is. In an intuitive sense, causality explains how different events can happen as the result of other sets of events or factors. For example, I noticed when I run fast, my heart rate increases. This is a correlation between my heart rate and running.

[1] GGGI: "The GGGI measures gender-based gaps in access to resources and opportunities rather than the actual levels of available resources and opportunities. he GGGI evaluates countries based on outcomes rather than inputs or means."(Pereznieto & Marcus, 2015)

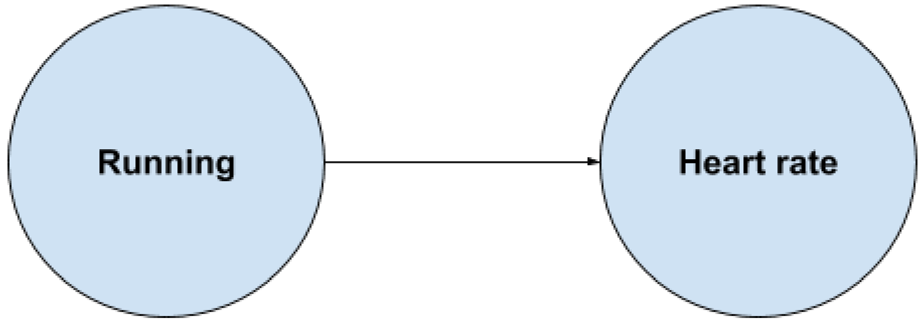

Figure 2. The graph of my association is based on observing two variables only.

Since running fast and increased heart rate are correlated, it does not mean that they cause each other. An increase in blood pressure does not happen in a vacuum. As seen in Figure 3, some factors affect both the heart rate and running. In this Figure, the amount of body fluid can be a common factor for both. We can also see that there is another association between having a history of heart disease and heart rate.

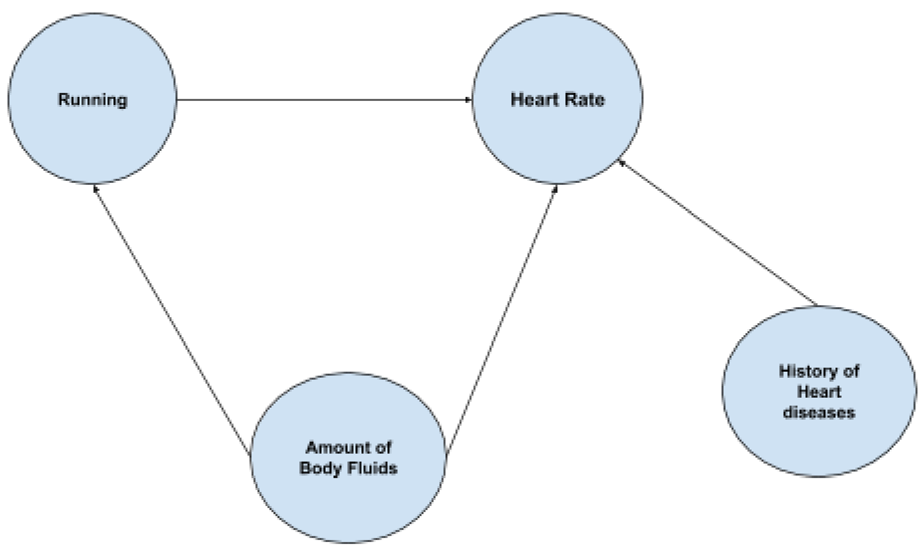

Figure 3. The graph of my association is based on more factors that happen in the real world.

To test if heart rate and running have a casual relationship, we would run a randomized control trial, RCT, to control for other factors that can affect heart rate and the running. For example, we can get a group that has the same amount of body fluids and the same history of heart disease. Then, we split the group into two, control and trial. The control group does not run, but the other group runs for 10 minutes. We measure the heart rate pre-running for both groups and measure the rate again after 10 minutes.

Table 1. Imaginary data collected to illustrate how RCTs are utilized.

	Mean Pre-experiment heart rate	Mean Post-experiment heart rate	Difference
Control	100 bpm	103 bpm	3 bpm
Treatment	120 bpm	230 bpm	110 bpm

The average heart rate is then computed for both groups and then subtracted to calculate the difference between the treatment and the group. We can see that there is a difference in the heart rate when controlled for some of the factors. This is not a complete example, to make this test more accurate, we need to incorporate p-value testing and add control for more variables such as the gender and the age to show statically significant results.

Conducting RCTs in the real world is challenging, i.e. we cannot control for the weather on a specific day or the level of Carbon Dioxide in the atmosphere. In other words, our association and correlations between events can not be proven casual using RCTs. So what is the alternative? Pearl's model for causation.

Pearl's Model

Pearl's model of causation is simply put as a "casual ladder." It involves three levels to establish a causal analysis: association, intervention, and counterfactuals (Pearl et al., 2016).

Association

As Pearl describes in his book, the association is looking for regularities among the data. Figure 1 for instance is an example of an association between GGGI and the percentage of women graduates in STEM. In other words, the association uses observations from data to formulate questions/hypotheses about the relationship between different variables (Pearl et al., 2016).

Intervention

It is the second level in Pearl's model, and it is concerned with what-if questions. For instance, what if all countries had the same GDP per capita, what is the effect of GGGI on the percentage of women graduates in STEM? In other words, we know we cannot run an RCT to control for confounding variables. However, we can control for the confounding variables using interventions.

Counterfactuals

The last step to test the robustness of the causal analysis to test if the outcome would have taken place if we got to observe some of the unobserved factors. For example, what would be the percentage of STEM female graduates if Rwanda's genocide didn't happen? This step is the highest in the causality ladder as we don't observe the counterfactual cases.

So, what went wrong in Figure 1?

Following Pearl's methodology, the first step was to see if there was an association between gender equality and the percentage of women graduating in STEM. First, I downloaded the same dataset, United Nations Educational, Scientific, and Cultural Organization (UNESCO) for 2015, using the World Bank DataBank (*Gender Statistics* | *DataBank*, n.d.). The sample had 77 countries. Second, instead of using the Global Gender Gap, I used the Basic Index of Gender Inequality (BIGI). GGGI does not count for STEM investments (Stoet & Geary, 2019). It does not account for a gender quota in the electoral system in its measures. Moreover, gender equality is a construct, and with the diverse set of indices that aim to measure the gender gap, it is important to replicate studies using different indices to test the robustness of the generalization. For instance, using the GGGI might show a significant correlation while using the BIGI shows no significance. Third, I only counted countries that had values for both BIGI and women's percentages in STEM graduates, which resulted in a sample size of 30 countries.

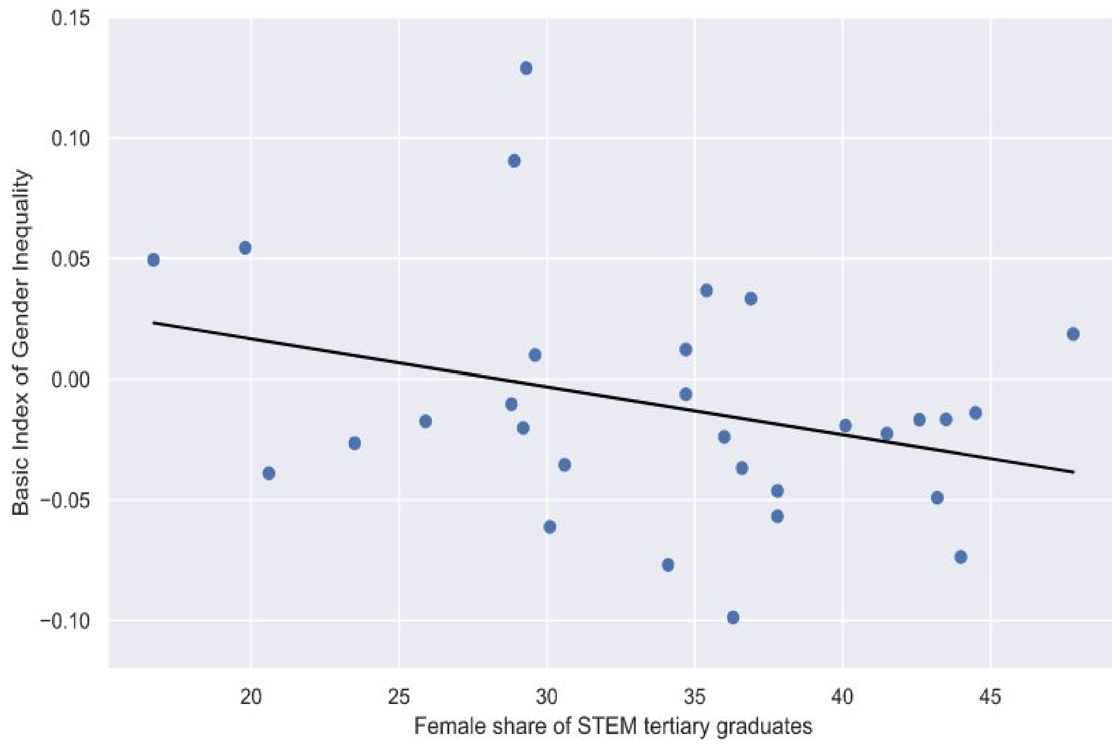

Figure 4. Replication of Figure 1 using a different index of gender equality. Source code and documentation are found here.

I conducted a Spearman Rank-Order Correlation to measure the strength of the monatomic relationship between the two variables. With a sample size of 30 countries, $r_2 = -0.257$ and with a $p = 0.170$. This shows a statistically non-significant correlation between the two variables. We cannot conclude that gender equality and women's shares of STEM graduates are causally related.

Pearl’s Framework in Action

Proposed Approach

We can conclude that using weak correlation, which is not significant, to imply causation is empirically flawed. To understand how gender equality affects STEM participation, it is important to incorporate other factors such as economic development, the sex ratio in the population, and cultural norms.

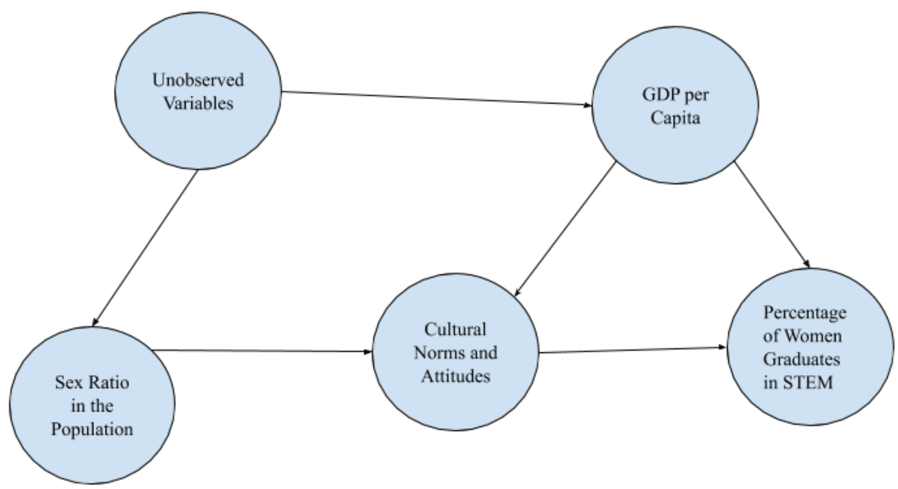

Figure 5. Directed Acyclic Graph for the new study proposal that investigates the relationship between gender norms, gender equality, and the percentage of women graduates in STEM.

The second step from the ladder, is an intervention, and by assigning a certain value to the GDP per capita, we can measure the causal effect of the treatment, the cultural norms, and the outcome of women shared in STEM graduates. By assigning a certain value for the GDP per capita, is what we call an intervention in the system. In other words, it means we are examining the relationship between the treatment and the outcome in a world where all GDP per capita is the same for all data points, known as the do-calculus. This intervention comes in handy as we adjust the casual graph to match the new structure.

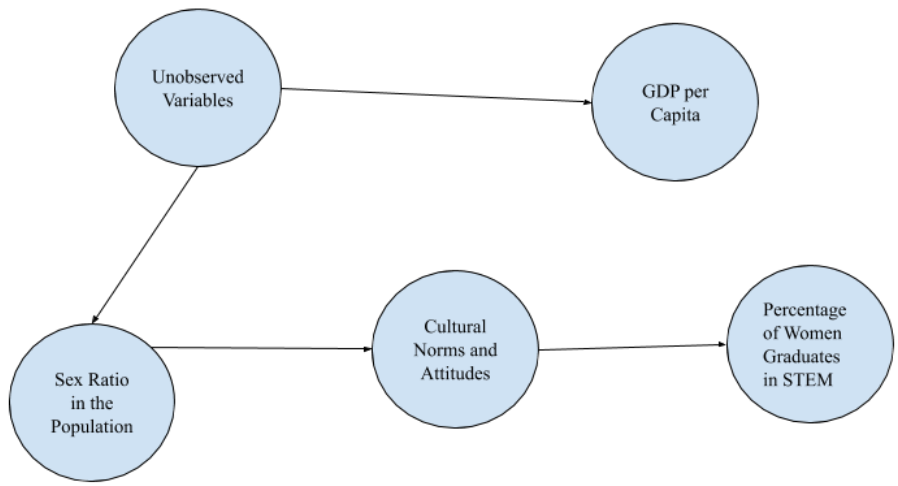

Figure 6. Directed Acyclic Graph updated after intervening on the GDP per Capita variable.

We used GDP per capita as it is a common cause for the treatment and the effect, hence it is a confounder we need to control for. Lastly, The counterfactual step, then, would calculate the new treatment effect of economic status on participation in STEM if we got to observe some of the unobserved variables for a specific country. For instance, what would be the percentage of women in STEM graduates in Rwanda if the Rwandan genocide of 1995 did not happen? As seen in the casual graph, the unobserved variable that affects both GDP and sex ratio, in the Rwandan case, is the genocide.

Key challenges

This proposal is not perfect, but it shows what it would take to have a robust causal analysis of norms and attitudes towards gender, like gender equality. One of the key challenges is to define and measure the norms and the social construct of gender equality. Using indices for that purpose can be misleading. For instance, changing the index from GGGI to BIGI showed statistically insignificant results that contradict Stoet and Geary's findings. An alternative to indices can be longitudinal surveys about gender. The second key challenge is the unobserved variables. We don't know what our world would look like if the interventions took another course of action based on a new set of information that we often don't see, like in a parallel universe. For example, we only observe post-genocide Rwanda, and its corresponding GDP, sex ratio, and percentage of women in STEM, but we don't observe the alternative Rwanda where the genocide did not happen, and we don't have its corresponding values for the variables of interest. Thus, we need to find alternatives to imagine that parallel universe, such as synthetic controls and matching. To make these alternatives work, we need to create variables to indicate whether a country had a genocide nor not, participated in World War I, etc. This process is also taxing, each country has its rich history and contextual factors. Thus, instead of investigating all 195 countries, narrowing the scope down to individual countries with similar historical contexts is more practical.

Conclusion

This piece shows the importance of looking at causality with a critical lens. Correlation does not mean causation. The statement is simple, and it is often overlooked. Pearl's framework allows us to analyze potential causal relationships utilizing directed acyclic graphs (DAGs). The framework is not enough. It is important to acknowledge and communicate the limitations of the methods and the data. Researchers need to also be critical and thorough in their studies especially when these studies are presented to the public.

Reference

Asmaaalaa99/casual_inference. (n.d.). GitHub. Retrieved October 18, 2020, from https://github.com/asmaaalaa99/casual_inference

Corrigendum: The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. (2020). *Psychological Science*, *31*(1), 110–111. https://doi.org/10.1177/0956797619892892

Gender Statistics | DataBank. (n.d.). Retrieved October 18, 2020, from https://databank.worldbank.org/source/gender-statistics

Khazan, O. (2018, February 18). *The More Gender Equality, the Fewer Women in STEM*. The Atlantic. https://www.theatlantic.com/science/archive/2018/02/the-more-gender-equality-the-fewer-women-in-stem/553592/

Kiersz, A. (n.d.). *Two male academics have proposed a new measure of gender inequality that suggests men are falling behind women in 91 countries*. Business Insider. Retrieved October 18, 2020, from https://www.businessinsider.com/basic-gender-inequality-index-methodology-2019-1

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Pereznieto, P., & Marcus, R. (2015). *What can gender indices tell us about gender norms that affect adolescent girls?* https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9813.pdf

SAGE PDF Full Text. (n.d.). Retrieved October 18, 2020, from https://journals.sagepub.com/doi/pdf/10.1177/0956797619892892

Snapshot. (n.d.). Retrieved October 18, 2020, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205349

Stoet, G., & Geary, D. C. (2019). A simplified approach to measuring national gender inequality. *PLOS ONE*, *14*(1), e0205349. https://doi.org/10.1371/journal.pone.0205349

The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education—Gijsbert Stoet, David C. Geary, 2018. (n.d.). Retrieved October 18, 2020, from https://journals.sagepub.com/doi/10.1177/0956797617741719