ME41125 Introduction to Engineering Research, Assignment: research paper
Brigitte de Graaf, Delft, The Netherlands, compiled June 5, 2021

Correlation between female independence and children that experience domestic violence - A data analysis

Brigitte de Graaf

***Abstract*—This paper aims to find a correlation between female independence and the number of children that experience violence. As indicator of female independence the gender wage gap and a variable representing divorce rates are used. The data is collected from various sources that have to uphold variable-specific standards. Complete data sets are sorted from all the found data. These are used to determine whether the sample size is sufficient, what the correlation is and what the multiple correlation coefficient is. Two out of three correlations contradict the expectation presented in the beginning of section II. Methods. A potential explanation for this is that the complete data sets are often from Europe which causes less developed regions not to be considered. The hypothesis "There is a correlation between financial independence of women, the divorce rate and children that experience violence globally.". is accepted because the multiple correlation coefficient shows a moderate degree of correlation.**

	Divorcing (D)		*High-Conflict Nondivorcing (HC-ND)*		*Low-Conflict Nondivorcing (LC-ND)*		*Significant Contrasts*
	Boy	*Girl*	*Boy*	*Girl*	*Boy*	*Girl*	
Externalizing	.47	.24	.49	.29	−.18	−.28	D,HC-ND > LC-ND B > G
Internalizing	−.03	.02	.23	.21	−.15	−.13	HC-ND > LC-ND
Cognitive agency	−.20	−.01	−.42	−.23	.04	.28	LC-ND > D > HC-ND G > B
Social responsibility	−.15	−.05	−.40	−.28	.02	.19	LC-ND > HC-ND G > B
Social competence	.01	.06	−.13	−.16	.19	.17	LC-ND > HC-ND
Self-esteem	.04	−.03	−.15	−.19	.20	.16	LC-ND > HC-ND

Fig. 1. Adjustment of children whose parents will later divorce and those in high- and low-conflict nondivorcing families [6]

I. Introduction

"Divorce has been shown to diminish a child's future competence in all areas of life, including family relationships, education, emotional well-being, and future earning power." [1]. Multiple studies show and it is widely accepted that divorce has a negative effect on a child's future. The study previously referred to was performed in the United States, which has a divorce rate among the highest in the world [2]. Often these studies are performed in countries with a welfare comparable to the United States. These countries are also classified as the best countries to raise kids [3]. This seems at first sight contradictory and therefore when trying to create a best possible environment for children to grow up in, questions are raised such as, is divorce always a bad thing for children?, when is it a good thing?, and what situation allows a divorce that benefits the well being of a child?

To answer these questions first the circumstances that can cause divorce are considered. This can be forms of abuse such as physical, psychological, sexual but also infidelity, addiction and adjustment problems can be factors that play a role [4]. Children notice the problems and can be effected by them. Indirectly, by experiencing the unhealthy environment created in the home, which has a variety of negative affects on children. Directly, because children with parents that use physical violence against each other are also at a higher risk of getting physically harmed themselves [5]. To answer the question of when is divorce a good thing for children the consequence of children living in low conflict nondivorcing home, a high conflict nondivorcing home and a divorced household are compared. The results, presented in figure 1, show that in high-conflict situations a divorce would be better for children [6].

All authors are with the institute TU Delft
j.m.degraaf@student.tudelft.nl

It is important to look at why divorces are more common in certain countries and why it was lower in history to answer the question, what situation allows a divorce that benefits the well being of a child? When thinking about the main factors that caused an increase in divorce rates one clear relation that can be found is that a higher share of a wife's income relates to a higher chance at divorce [7]. This is very intuitive since the financial means a person has available determines a persons entire lives, whether the person can eat and drink, has a home etc.. Being financially independent from someone else will result in a person being less likely to leave that person, even in an unhealthy situation. The financial independence of women can be assessed by looking at e.g. the gender wage gap.

So far, research has been done towards the negative effects of a divorce on children are done in wealthier countries [8], [9], [10]. However, research also showed that in cases of High-conflict nondivorcing families the child actually developed worse than a child in a divorced family [6]. Since a higher divorce rate can be linked to the (financial) independence of women the question that this paper aims to answer is "To which extent is there a correlation between the financial independence of women, the divorce rate and children that experience violence globally?" The answer to this question will give more information on how to most efficiently target social issues, potentially multiple at the same time if turns out that an increase in financial independence of women relates to a decrease in children that experience violence.

II. Methods

To find an answer to this question hypotheses are made. The null hypothesis is defined as "There is no correlation between financial independence of women, the divorce rate and children that experience violence globally.". The alternative hypothesis states "There is a correlation between financial

Country	Population	Continent	UN Statistical subregion
-	amount	-	-

CDR	CMR	gender wage gap
per 1000 inhabitants	per 1000 inhabitants	%
%		

Children that experienced violent discipline
%

TABLE I
CHARACTERISTICS OF THE STUDY SAMPLE, THE FIRST ROW SHOWS WHAT VALUE, THE SECOND ROW REPRESENTING THE UNIT.

independence of women, the divorce rate and children that experience violence globally.". To find out if the alternative hypothesis is correct data is gathered for all three categories. This data is then pre-processed. Finally, the data is analysed by looking at correlation and multiple correlation. The prepossessing and data processing has been performed using *MATLAB R2021a* and *Microsoft Excel for Microsoft 365 MSO (16.0.14026.20202) 32-bit*.

A. Data collection

To collect the data various standards were upheld to validate the correctness of the data. With the focus on finding correlations that were as accurate as possible, it was tried to gain the data from the smallest amount of different sources as possible whilst still upholding certain standards and criteria. The standard used for all the data was that it had to come from international humanitarian organizations or websites of governments. When multiple sources were available a trade-off between the amount of available information and the size of the humanitarian organisation or reliability of the government that published it was made. The manner in which all the data of these different categories is stored in is visible in table I.

1) Crude marriage rate and Crude divorce rate: The source used for the crude marriage rate (CMR) and the crude divorce rate (CDR) is from the United Nations (UN) Department of Economic and Social Affairs [11]. The reason CMR and CDR are chosen as measures is because this represents the number of marriages and the number of divorces per 1000 people in a specific country. Because it is averaged over a 1000 people a comparison between countries can be made. By looking at both the CMR and the CDR, instead of solely the divorce rate, an estimation for the percentage of divorces per marriages is made. The further requirements for the data is that the data from the latest possible statistics and only statistics based on 2000 AD or later are considered.

2) Gender wage gap: For the data of the gender wage gap two different sources were used since one source simply did not provide enough data. The source used is [12] complemented by [13]. This decision has been made because *EUROSTAT* had more statistics which makes the standard deviation of the data more consistent. The amount statistics was the decisive factor since the validity of both institutions is comparable, *EUROSTAT* is the statistical office of the European Union [14] and *OECD* is an intergovernmental economic organisation that consists of 37 member countries [15].

3) Children that experience violence: The final category has one main source and a variety of different sources to back up these statistics and complement where possible. This main source used is a report from *Known violence in childhood* [16]. This data is validated and complemented using various sources among which research published by *UNICEF* both through the database [17] and a report [18]. It is important to note that the report describes various studies that have been performed in various countries. This data is therefore not immediately comparable. To be able to still use part of this data and data from other sources criteria that the data needs to fulfill were drawn up:

- The study has been performed not earlier then the year 2000.
- The study covers multiple year layers, preferably 2-14 years old.
- The study considers all form of physical domestic violence.
- The study classifies children that have experienced any physical violence on at least one occasion. The minimum number of occasions for the child to be classified as "child that experienced physical violence by heir caretakers" cannot be higher then one. (Neither can the classification be based on (a) specific form(s) of physical violence instead of covering all.)

Other sources used to gather the data were mainly government websites [19], [20], [21], [22], [23], [24], [25], [26], [27].

4) Population, continent and UN Statistical subregion: To determine global trends it is important to mathematically weight the data from the various countries. This is done by the population size, the bigger a population the larger the contribution of the data acquired from that country is. Besides dividing the data in the category countries, the countries are divided based on continent and UN Statistical subregion. The decision to also look into regions and continents instead of simply looking at countries is made because of the expectation that this way, the results draw a broader picture and certain trends in the data will allow for better interpretation and determining of the reasons that cause these trends. How countries are divided over continents is common knowledge, however UN Statistical subregions will be further explained. UN Statistical subregions are regions that consist of countries that are classified as that region based on their location, income group and subregion of the world. The various subregions are *Eastern*, *Middle*, *Northern* and *Western Africa*, *Easter*, *South-Central*, *South-Eastern* and *Western Asia*, *Eastern*, *Northern*, *Southern* and *Western Europe*, *Latin America and the Caribbean*, *Northern America*, *Oceania*, *Sub-Saharan Africa* and *Least developed countries* [28].

B. Data pre-processing

The incomplete data sets in the data acquired need to be filtered out. The data from the various sources is joined in one *Excel* sheet. This sheet includes various columns; country, population (amount), Continent, UN Statistical subregion, CDR (number of divorces per 1000 inhabitants), CMR (number of marriages per 1000 inhabitants), gender wage gap (%), children that experienced violent discipline during childhood (%). The found data for the last four categories are all inserted

in the file. This results in some countries not having a complete data set. For this reason, the countries with an incomplete data set are filtered out. To filter out these countries a matrix with all the data is made. Then the *NaN* values in the matrix are located. The rows with *NaN* values have to be deleted to determine the final matrix with solely complete data sets. Because for some countries more than one value is missing this results in the index of this row occurring multiple times in the array with to be deleted rows. An array is made that contains the row numbers that should be deleted. Finally, these rows are deleted from the matrix.

C. Data processing

1) CMR and CDR: To make the number used to represent the amount divorces more generic the decision was made to divide the CDR by the CMR. This new number multiplied by ten represents the percentage of marriages that end in divorce.

2) Weighing the complete data sets: To determine the correlation between the values the complete data sets are weighted by the population that it relates to. However, since weighting it by the exact numbers is not possible in *MATLAB* because this takes too much storage the population is rounded to 1000s. The amount of population is then divided by 1000 and this number is used as the weight by copying the row containing the other data from that country the number of times as the weight. Doing this for all countries results in the weighted matrix. A visual representation of this data is visible in figure 2

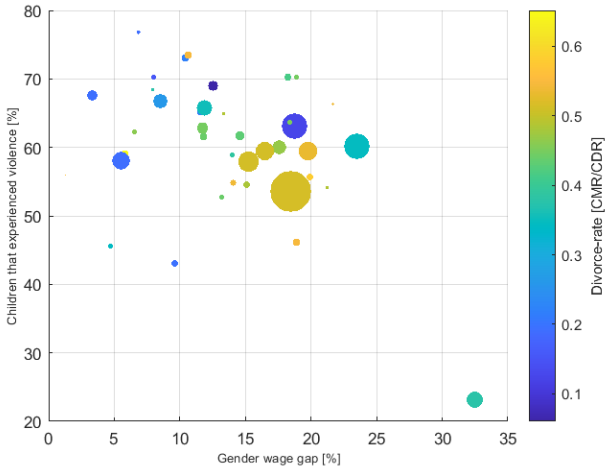

Fig. 2. Scatterplot of the gender wage gap in [%] on the x-axis, Children that experienced violence in [%] on the y-axis, the divorce-rate represented by the color and the size of the point corresponds to the size of the population that the data point relates to.

3) Sample size: To determine whether the sample analysed is enough to scale up the results to population level first the sample size is determined. This is done by adding up the population in the countries with complete data sets. This value is then divided by the total global population [29]. This results in equation 1.

$$\frac{sample\ size}{total\ global\ population} = \frac{1.258.874.580}{7.674.000.000} = 16.4\ \% \quad (1)$$

The confidence level desired is 95 %. The z-value corresponding to this confidence interval is 1.96. The margin of error MOE chosen is 5 %. To determine the number of samples n necessary equation 2 is used [30]. The σ in this equation is calculated by averaging the variable with the biggest standard deviation.

$$n \geq \left(\frac{z^{\star}\sigma}{MOE}\right)^2 \quad (2)$$

Applying this results in $n = 119.629.000$. Since in the analysis performed $n = 1.258.874.580$ this is sufficient.

4) Correlations: First three correlations are determined, between divorce-rate and gender wage gap, between gender wage gap and children that experience violence, and finally between the divorce-rate and children that experience violence. The value of the correlation can be classified as perfect (± 1), high degree (between ± 0.5 and ± 1), moderate degree (between ± 0.3 and ± 0.49), low degree (between ± 0.29 and ± 0) and no correlation (0) [31]. The correlations are determined using the *crosscoef* command in *MATLAB*.

5) Multiple Correlations: To find the correlation between the three variables the formula expressed in equation 3 is used. In this formula, x represents the ratio $\frac{CDR}{CMR}$, y represents the gender wage gap (%), z represents the percentage of children that experienced violence, r represents the correlation calculated previously and R is represents the multiple correlation [32].

$$R_{z,xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xz}r_{yz}r_{xy}}{1 - r_{xy}^2}} \quad (3)$$

Since $R_{z,xy}$ is a biased estimate of the population multiple correlation coefficient. However using equation 4 this bias can be reduced. In this equation R_{adj} represents the less biased multiple correlation, n represents the number of data elements in the sample and k the number of variables [32].

$$R_{adj}^2 = 1 - \frac{\left(1 - R_{z,xy}^2\right)(n-1)}{n-k-1} \quad (4)$$

The value of the coefficient of the multiple correlation indicates how well the variable can be predicted by the linear function. The value is somewhere in between 0 and 1. The higher it is the better it can be predicted.

III. Results

A. Correlation

The correlations found and their interpretations are shown below.

- between divorce-rate and gender wage gap: 0.2077. This can be classified as a low degree of correlation. This means that there is a small correlation. Because it is positive it means that with an increase in divorce-rate the gender wage gap also increases.
- between gender wage gap and children that experience violence: -0.5871. This can be classified as a high degree of correlation. This means that there is a strong correlation. Because it is negative it is interpreted as an increase

in the gender wage gap results in a decrease in children that experience violence.

- between the divorce-rate and children that experience violence: -0.2912, This can be classified as a low degree of correlation. This means that there is a small correlation. Because it is negative it is interpreted as an increase in divorce-rate results in a decrease in children that experience violence.

B. Multiple Correlation

The biased value results in $R_{z,xy} = 0.6121$ using equation 3, when using equation 4 the result is $R_{adj} = 0.3747$. This indicates a moderate degree of correlation.

IV. Discussion

The sample size of the data is large enough to scale it up when taking into account an error margin of 5 %. There was a negative correlation between the gender wage gap and children that experience violence, contradictory to what was expected based on the literature study. Likewise, the correlation between divorce-rate and gender wage gap showed contradictory results. A reason for this could be that the divorce rate and the gender wage gap are both used as indicators for female independence. However social topics like these are not solely dependent on each other and many more factors play a role. To reduce the effect of the other factors this study looked at the numbers of the variables globally. However, the complete data sets were mainly from European countries. This becomes clear when looking at figure 3, in which the orange bars represent the number of countries from a specific continent used in the analysis whilst blue is the number of countries actually in that continent [33].

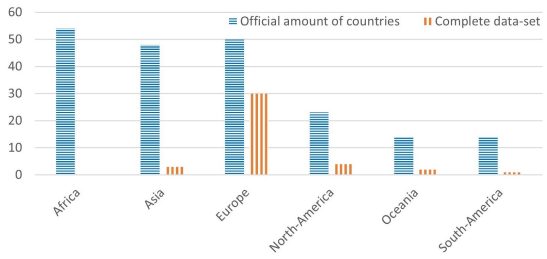

Fig. 3. Visual representation of which the countries that are used to do the analysis compared to the number of countries in the continent.

This causes less developed regions not to be considered whilst these are the regions that generally have the most children that experience violence [16]. The division over regions of the number of children that experience violence is shown in figure 4.

Besides increasing the amount of data on the gender wage gap, which is the main cause for the distribution visible in figure 3, the data would be more reliable if it all came from one source. Ideally, all variables, but the accuracy of this analysis would improve if one large source provided all data necessary for (at least) one variable. The reason for this is that in some cases both sources gave a value for a country but the value varied substantially whilst both sources were from

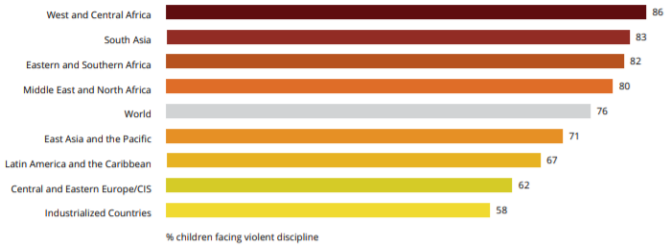

Fig. 4. Corporal punishment at home (children aged 1-14) by region, 2015. [16]

proper sources. This is an issue that can only be resolved by using one source since number social studies often have a lower accuracy but high precision. Another thing to note is that CMR and CDR are studies from no later than the year 2000 whilst the report used for the percentages of children that experienced violence is from the year 2017. Therefore the CMR and CDR values do not completely overlap. However, because the research is focused on finding trends, as long as all the data has the same bias the difference in the results is expected to be minimal since major changes in the ratios between the data from countries in such a short amount of time is deemed unlikely. Finally, the study is framed in such a way that it does not consider children born out of wedlock.

V. Conclusion

The data shows that, the correlation between the divorce-rate and gender wage gap has a low degree of positive correlation, the gender wage gap and percentage of children that experienced violence have a high degree of negative correlation and finally, the percentage of children that experienced violence and the divorce-rate show a low degree of negative correlation. The small correlation found between the divorce-rate and the children that experience violence is in line with the argument made in the introduction, that divorce can also have positive effects on children. However, the other correlations go against the expectation presented in the introduction. The sample size is sufficient to scale from the sample to the population when the MOE is chosen as 5 % the results of the analysis performed on the sample can therefore be applied. These results are contradictory to the expectation in the introduction based on literature review and suggest that a higher gender wage gap has a positive effect on children and that a decrease in gender wage gap results in an increase in divorces. Based on the fact that the gender wage gap and divorce rate are used as indicators for female independence the data analysed in this paper would suggest that an increase in independence for women relates to an increase in children that experience violence. Besides this, the hypothesis can be accepted because when calculating the coefficient of multiple correlation a moderate degree of correlation was found.

VI. Data Availability Statement

The data analysis scripts/code that are used in this study are available in the corresponding GitHub repository de Graaf (2021) analysis_IER [Source code]. https://github.com/BrigittedeGraaf/IER.git.

ACKNOWLEDGMENTS

I want to thank Heike Vallery, Angeniet Kam and Lisa Hoogendam for providing the template document for this paper and the guidance throughout the process of writing it. I also to thank Yan Aung for his knowledge of the public sector and for advising me on how to approach this kind of topic. Finally, I would like to thank the students that reviewed this paper for their future feedback.

REFERENCES

[1] J. Anderson, "The impact of family structure on the health of children: Effects of divorce," *The Linacre Quarterly*, vol. 81, no. 4, pp. 378–387, 2014.

[2] C. Halligan, I. J. Chang, and D. Knox, "Positive effects of parental divorce on undergraduates," *Journal of Divorce & Remarriage*, vol. 55, no. 7, pp. 557–567, 2014.

[3] U.S.News. (2020) Best countries for raising kids. [Online]. Available: urlhttps://www.usnews.com/news/best-countries/best-countries-to-raise-a-family

[4] B. Vasudevan, D. M. Geetha, A. Bhaskar, B. Areekal, A. Lucas *et al.*, "Causes of divorce: a descriptive study from central kerala," *Journal of evolution of medical and dental sciences*, vol. 4, no. 20, pp. 3418–3427, 2015.

[5] G. Hornor, "Domestic violence and children," *Journal of Pediatric Health Care*, vol. 19, no. 4, pp. 206–212, 2005.

[6] E. M. Hetherington, *Coping with divorce, single parenting, and remarriage: A risk and resiliency perspective*. Psychology Press, 2014.

[7] G. Liu, A. Vikat *et al.*, "Does divorce risk depend on spouses´ relative income? a register-based study of first marriages in sweden in 1981–1998," Max Planck Institute for Demographic Research, Rostock, Germany, Tech. Rep., 2004.

[8] P. R. Amato and B. Keith, "Parental divorce and adult well-being: A meta-analysis," *Journal of Marriage and the Family*, pp. 43–58, 1991.

[9] E. M. Hetherington, "Social support and the adjustment of children in divorced and remarried families," *Childhood*, vol. 10, no. 2, pp. 217–236, 2003.

[10] A. J. Cherlin, P. L. Chase-Lansdale, and C. McRae, "Effects of parental divorce on mental health throughout the life course," *American Sociological Review*, pp. 239–249, 1998.

[11] World marriage data. [Online]. Available: https://www.un.org/en/development/desa/population/publications/dataset/marriage/data.asp

[12] Gender pay gap statistics. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics

[13] Gender wage gap. [Online]. Available: https://data.oecd.org/earnwage/gender-wage-gap.htm#indicator-chart

[14] Who we are. [Online]. Available: https://ec.europa.eu/eurostat/web/main/about/who-we-are

[15] Our global reach. [Online]. Available: https://www.oecd.org/about/members-and-partners/

[16] S. Kumar. (2018) Global report 2017 - ending violence in childhood. [Online]. Available: http://globalreport.knowviolenceinchildhood.org/

[17] Violent discipline. [Online]. Available: https://data.unicef.org/topic/child-protection/violence/violent-discipline/

[18] Violence against children in europe. [Online]. Available: https://www.unicef-irc.org/publications/pdf/violence_against.pdf

[19] (2019, may) Characteristics and outcomes of childhood abuse. [Online]. Available: https://www.abs.gov.au/articles/characteristics-and-outcomes-childhood-abuse#:~:text=of%20childhood%20abuse.-,How%20many%20Australian%20adults%20have%20experienced%20childhood%20abuse%3F,who%20experienced%20childhood%20sexual%20abuse.

[20] (2020, april) Children exposed to family violence. [Online]. Available: https://www.aihw.gov.au/reports/children-youth/australias-children/contents/justice-and-safety/children-exposed-to-family-violence

[21] Unicef reports high levels of child abuse in hungary. [Online]. Available: https://english.radio.cz/unicef-reports-high-levels-child-abuse-hungary-8078594#:~:text=The%20study%20shows%20that%20about,because%20of%20abuse%20or%20maltreatment.

[22] Multi-country study on the driversof violence affecting children. [Online]. Available: https://www.istitutodeglinnocenti.it/sites/default/files/drivers_of_violence_italy_0.pdf

[23] United nations study on violence against children. [Online]. Available: https://www.ohchr.org/Documents/HRBodies/CRC/StudyViolenceChildren/Responses/Netherlands.pdf

[24] S. Hillis, J. Mercy, A. Amobi, and H. Kress. (2016) Global prevalence of past-year violence against children: a systematic review and minimum estimates.

[25] Ending violence against children. [Online]. Available: https://www.norway.no/en/missions/wto-un/our-priorities/humanitarian-affairs/ending-violence-against-children2/

[26] (2018) Survey of adverse childhood experiences and associated health-harming behaviours among polish students (2018). [Online]. Available: https://www.euro.who.int/en/countries/poland/publications/survey-of-adverse-childhood-experiences-and-associated-health-harming-behaviours

[27] (2018) Swissinfo. [Online]. Available: https://www.swissinfo.ch/eng/corporal-punishment-_violence-in-childhood---also-a-swiss-problem/43684526

[28] (2017) Population division. [Online]. Available: https://web.archive.org/web/20181111202501/https://esa.un.org/unpd/wpp/General/Files/Definition_of_Regions.pdf

[29] (2019) Population, total. [Online]. Available: https://data.worldbank.org/indicator/SP.POP.TOTL

[30] D. J. Rumsey. How to determine the minimum size needed for a statistical sample. [Online]. Available: https://www.dummies.com/education/math/statistics/how-to-determine-the-minimum-size-needed-for-a-statistical-sample/

[31] C. Zaiontz. Pearsonâs correlation coefficient. [Online]. Available: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/

[32] ——. Multiple correlation. [Online]. Available: https://www.real-statistics.com/correlation/multiple-correlation/

[33] List of countries by continents. [Online]. Available: https://simple.wikipedia.org/wiki/List_of_countries_by_continents