Global Families Project

Global Families Project Team

6/19/23

Table of contents

1 Project Summary

Gender inequality perpetuates harmful norms that justify violence against women and children and is associated with higher rates of family violence.

Worldwide, parental physical abuse is a common form of family violence that children are exposed to at alarming rates. Parental engagement in physical abuse is linked to negative child outcomes including depression, anxiety, and aggression that may persist into adulthood. Globally, these continuing mental health and aggression problems may have high financial costs, with effects both on social service systems and developing economies.

Despite the substantial scholarship on parent- and family-level predictors of parent-to-child physical violence, important questions remain about societal-level predictors of parental physical abuse and its associations with young children's development in developing and transitional countries.

A further gap in prior literature is the lack of studies that have examined potential moderators such as child age and household economic status in the associations between gender inequality and parental violence against children.

Using data from over 520,000 families in 57 low- and middle-income countries (LMICs), the current project seeks to address these research gaps by examining the associations of country-level gender inequality and violent social contexts with caregivers' use of physically abusive behavior and child social-emotional development. We will employ multilevel models using data on parental physical violence against children, family socio-economic characteristics, and children's social-emotional development from the UNICEF Multiple Indicator Cluster Surveys (MICS) and data on country-level gender inequality and violent social contexts from the United Nations Development Programme on Human Development and the World Health Organization Global Health Observatory.

The specific aims are to 1) examine the associations of gender inequality with parental child physical abuse in LMICs, and the moderating roles of child age and household economic status in these associations, 2) examine the associations of violent social norms and crimes with parental physical abuse in LMICs, and 3) examine the associations of parental physical abuse with child social-emotional development in the context of gender inequality and violent norms and crimes in LMICs, and whether country-level normativeness of physical abuse moderates these associations.

The proposed studies will advance the understanding of macro-level social and economic indicators that perpetuate caregivers' physical violence against children in international contexts.

Study findings will inform cross-cultural programs and policies that reduce gender disparities and prevent parental physical abuse to promote child social-emotional development across the globe.

In addition, these studies will provide rigorous research engagement opportunities to undergraduate students and graduate students and strengthen the research environment at the University of Michigan-Flint.

2 Research Team

Julie Ma, Principal Investigator

Associate Professor of Social Work, UM-Flint

Professor Ma's research interests center around the effects of neighborhood disadvantage and negative parenting on the well-being of children. Her research builds on her experience in parent education programs that serve families in marginalized communities in Michigan. Much of her current research focuses on the risks of negative contextual and family influences such as neighborhood poverty and disorganization, and parental corporal punishment on behavior problems and maltreatment in early childhood.

Andy Grogan-Kaylor, Co-Investigator

Sandra K. Danziger Collegiate Professor, Professor of Social Work

Professor Grogan-Kaylor's research focuses on knowledge development and intervention research on children and families with the aim of reducing violence against children and improving family and child wellbeing. Grogan-Kaylor's current research projects examine parenting behaviors such as physical punishment and parental expressions of emotional warmth and support, and their effects on children's aggression, antisocial behavior, anxiety, and depression.

Shawna Lee, Co-Investigator

Professor of Social Work

Professor Lee is a professor at the University of Michigan School of Social Work. She is the director of the Parenting in Context Research Lab and the director of the Program Evaluation Group at the School. Lee has published on topics related to child maltreatment, fathers' parenting, father-child relationships, parenting stress and family functioning, and parental discipline. Her recent research focuses on parenting and stress during the COVID-19 pandemic.

3 A Quick Introduction to R

3.1 Why?

R has a reputation for being difficult to learn, and a lot of that reputation is deserved. However, it is possible to teach R in an accessible way, and **a little bit of R can take you a long way**.

R is open source, and therefore free, statistical software that is particularly good at obtaining, analyzing and visualizing data.

R Commands are stored in a *script* or *code* file that usually ends in .R, e.g. `myscript.R`. The command file is distinct from your actual data, stored in an .RData file, e.g. `mydata.RData`.

A great deal of data analysis and visualization involves the same core set of steps.

Given the fact that we often want to apply the same core set of tasks to new questions and new data, there are ways to overcome the steep learning curve and learn a replicable set of commands that can be applied to problem after problem. **The same 5 to 10 lines of R code can often be tweaked over and over again for multiple projects.**

$$\text{have a question} \rightarrow \text{get data} \rightarrow \text{process and clean data} \rightarrow$$

$$\text{visualize data} \rightarrow \text{analyze data} \rightarrow \text{make conclusions}$$

3.2 Get R

R is available at https://www.r-project.org/. R is a lot easier to run if you run it from RStudio, http://www.rstudio.com.

3.3 Get Data

Data often comes from other types of data files like SPSS, Stata, or Excel. Especially in beginning R programming, getting the data into R can be the most complicated part of your program.

```r
load("the/path/to/mydata.Rdata") # data in R format

library(haven) # library for importing data
mydata <- read_sav("the/path/to/mySPSSfile.sav") # SPSS
mydata <- read_dta("the/path/to/myStatafile.dta") # Stata

library(readxl) # library for importing Excel files
mydata <- read_excel("the/path/to/mySpreadsheet.xls")

save(mydata, file = "mydata.RData") # save in R format
```

3.4 Process and Clean Data

The `$` sign is a kind of "connector". `mydata$x` means: "The variable `x` in the dataset called `mydata`".

```r
mydata$x[mydata$x == -9] <- NA # missing to NA
```

R makes a strong distinction between *continuous numeric* variables that measure scales like mental health or neighborhood safety, and *categorical factor variables* that measure non-ordered categories like religious identity or gender identity.

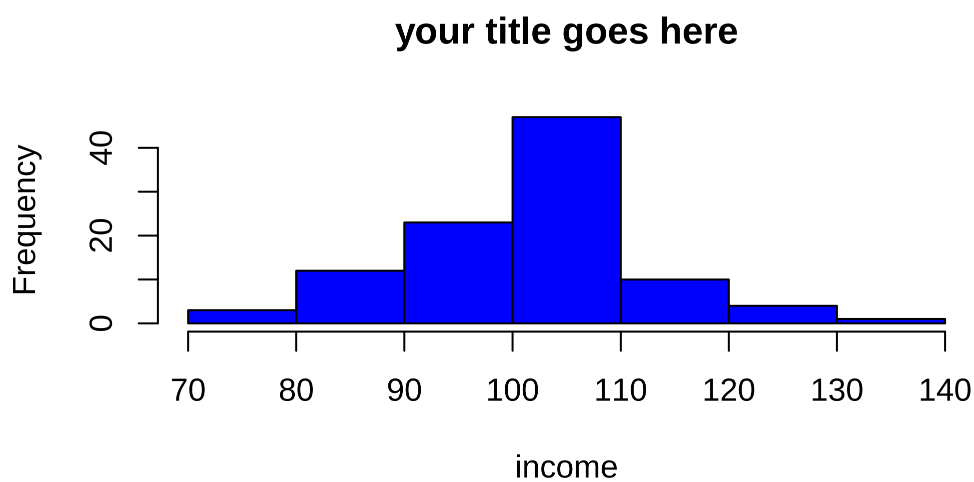
Many statistical and graphical procedures are designed to recognize and work with different variable types. You often *don't* need to use all of the options. e.g. `mydata$w <- factor(mydata$z)` will often work just fine. **Changing variables from factor to numeric, and vice versa can sometimes be the simple solution that solves a lot of problems when you are trying to graph your variables.**

```r
mydata$w <- factor(mydata$z, # original numeric variable
                   levels = c(0, 1, 2),
                   labels = c("Group A", "Group B", "Group C"),
                   ordered = TRUE) # whether order matters

mydata$z <- as.numeric(mydata$w) # factor to numeric
```

3.5 Visualize Data

3.5.1 Histogram

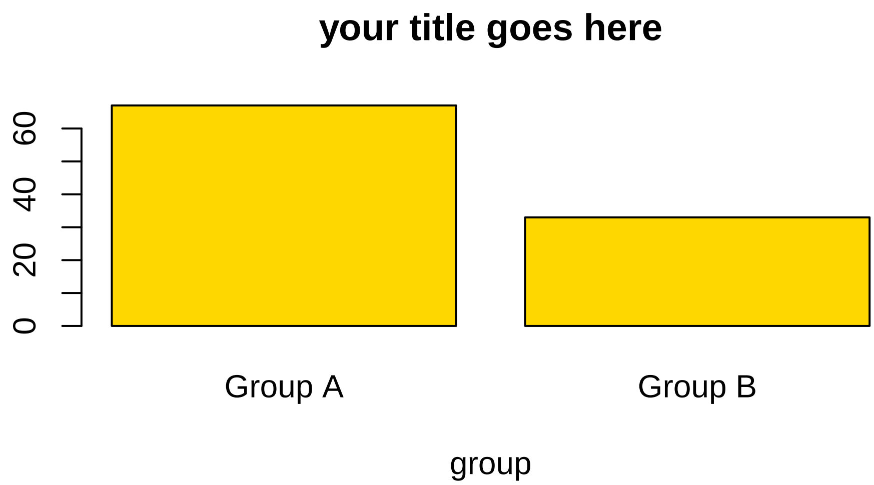
```
hist(mydata$x, # what I'm graphing
     main = "your title goes here", # title
     xlab = "income", # label for x axis
     col = "blue") # color
```



> **Tip**
>
> You often *don't* need to use all of the options. e.g. `hist(mydata$x)` will work just fine.

3.5.2 Barplot

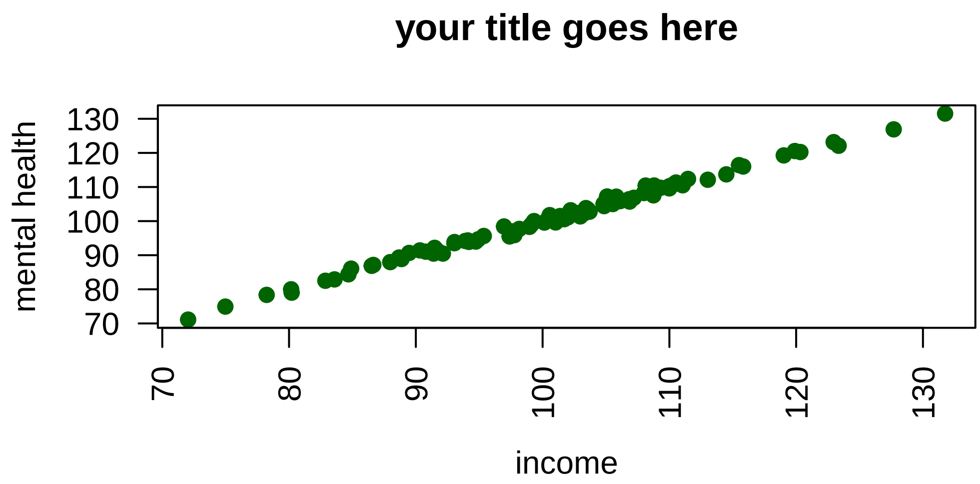
```
barplot(table(mydata$z), # what I'm graphing
        names.arg = c("Group A", "Group B"), # names
        main = "your title goes here", # title
        xlab = "group", # label for x axis
        col = "gold") # color
```



> **Tip**
>
> You often *don't* need to use all of the options. e.g. `barplot(table(mydata$z))` will work just fine.

3.5.3 Scatterplot

```
plot(mydata$x, mydata$y, # plot x and y
     main = "your title goes here", # title
     xlab = "income", # label for x axis
     ylab = "mental health", # label for y axis
     pch = 19, # Plot CHaracter, 19 is filled dots
     las = 2, # LAbel Style, 2 is "perpendicular"
     col = "darkgreen") # color
```



Tip

You often *don't* need to use all of the options. e.g. `plot(mydata$x, mydata$y)` will work just fine.

Tip

When scatterplots have fewer dots than you think they should have, often due to "overprinting", adding some random noise, or "jittering" the dots in the scatterplot may help: `plot(jitter(mydata$y, factor = 5000) ~ mydata$x)`. Experiment with different sizes of *factor*.

3.6 Analyze Data: Descriptive Statistics

```
summary(mydata$x) # for continuous or factor variables
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 72.04   94.05  102.04  101.16  108.16  131.75 
```

```
table(mydata$z) # especially suitable for factor variables
```

```
 1  2
67 33
```

For another approach to summarizing your data, try:

```
library(skimr)

skim(mydata)

skim(mydata$x)
```

References

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE Publications. https://doi.org/10.4135/9781849209366

Luke, D. (2004). *Multilevel modeling*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985147

Rabe-Hesketh, S., & Skrondal, A. (2022). Multilevel and longitudinal modeling using Stata. In *Stata Press* (4th ed.). Stata Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (pp. xxiv, 485 p.). Sage Publications.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis : Modeling change and event occurrence. In *Applied longitudinal data analysis : modeling change and event occurrence*. Oxford University Press.

A Simulating MICS Data

This appendix details the process of creating the simulated MICS data that is employed in the examples on this website.

MICS data are freely available, but usage of MICS requires completing a user agreement, and registering for a user account, on the MICS website, and thus MICS data should not be shared openly on a public website.

This Appendix is highly technical. It is not necessary to understand this Appendix to benefit from the rest of this website. However, the details of creating this simulated data may be of interest to some users.

A.1 Call Relevant Libraries

We need to call a number of relevant R libraries to simulate the data.

```r
library(tibble) # new dataframes

library(ggplot2) # nifty graphs

library(labelled) # labels

library(haven) # write Stata

library(tidyr) # tidy data

library(dplyr) # wrangle data

library(lme4) # multilevel models

library(sjPlot) # nice tables
```

A.2 Setup Some Basic Parameters of the Data

Because simulation is a random process, we set a *random seed* so that the simulation produces the same data set each time it is run.

We are going to simulate data with 30 countries, and 100 individuals per country.

```
set.seed(1234) # random seed

N_countries <- 30 # number of countries

N <- 100 # sample size / country
```

A.3 Simulate Data Based on MICS

This is multilevel data where individuals are nested, or clustered, inside countries. Excellent technical and pedagogical discussions of multilevel models can be found in Raudenbush & Bryk (2002), Singer & Willett (2003), Rabe-Hesketh & Skrondal (2022), Luke (2004), and Kreft & de Leeuw (1998).

A.3.1 Level 2

Simulating the second level of the data is relatively easy. We simply need to provide the number of countries, and then generate random effects for each country. Random effects are discussed in the above references, but essentially represent country level differences in the data.

```
country <- seq(1:N_countries) # sequence 1 to 30

u0 <- rnorm(N_countries, 0, .25) # random intercept

u1 <- rnorm(N_countries, 0, .05) # random slope

randomeffects <- data.frame(country, u0, u1) # dataframe of random effects
```

A.3.2 Level 1

Simulating the Level 1 data is more complex.

We `uncount` the data by 100 to create 100 observations for each country. We then create an `id` number.

We create randomly simulated parental discipline variables with proportions similar to those in MICS.

Lastly, we need to create the dependent variable. Because this is a dichotomous outcome, the process is somewhat complex. We need to craete a linear combination `z`, using regression weights derived from MICS. We then calculate predicted probabilities, and lastly generate a dichotomous `aggression` outcome from those probabilities.

```
MICSsimulated <- randomeffects %>%
  uncount(N) %>% # N individuals / country
  mutate(id = row_number()) %>% # unique id
  mutate(cd1 = rbinom(N * N_countries, 1, .38), # spank
         cd2 = rbinom(N * N_countries, 1, .05), # beat
         cd3 = rbinom(N * N_countries, 1, .64), # shout
         cd4 = rbinom(N * N_countries, 1, .78)) %>% # explain
  mutate(z = .7 + # linear combination based on MICS
           .23 * cd1 +
           .52 * cd2 +
           .42 * cd3 +
           -.21 * cd4 +
           u0) %>%
  mutate(p = exp(z) / (1 + exp(z))) %>% # probability
  mutate(aggression = rbinom(N * N_countries, 1, p)) %>% # binomial y
  select(id, country,
         cd1, cd2, cd3, cd4,
         aggression)
```

A.3.3 Variable Labels

We add variable labels to the data which will help us to understand the data as we analyze it.

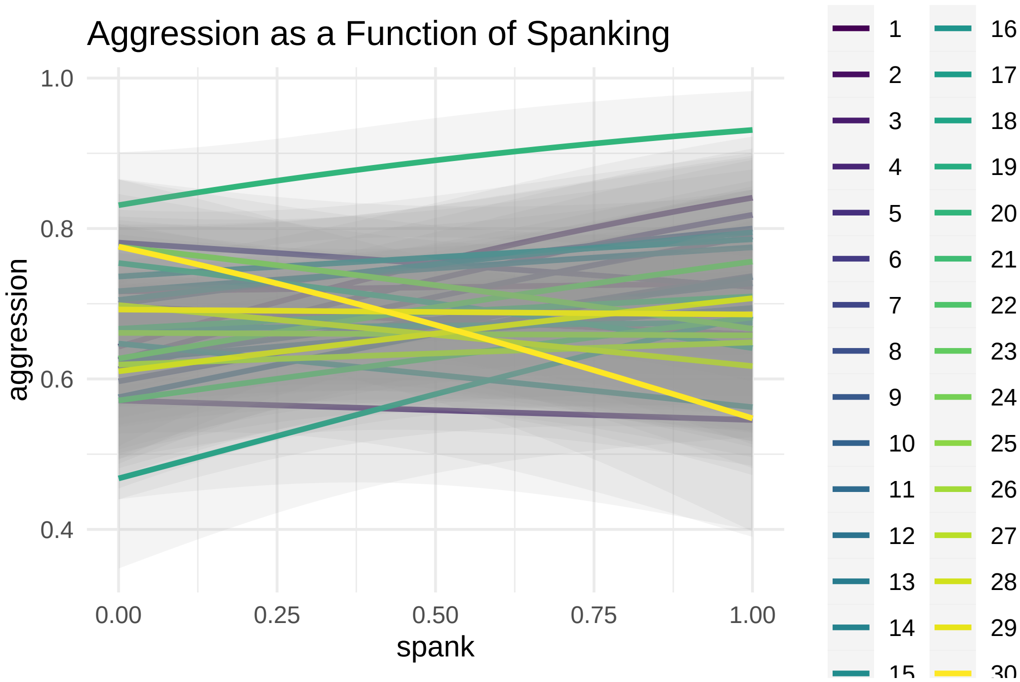
```
var_label(MICSsimulated$id) <- "id"

var_label(MICSsimulated$country) <- "country"

var_label(MICSsimulated$cd1) <- "spank"

var_label(MICSsimulated$cd2) <- "beat"
```

```r
var_label(MICSsimulated$cd3) <- "shout"

var_label(MICSsimulated$cd4) <- "explain"

var_label(MICSsimulated$aggression) <- "aggression"
```

A.4 Explore The Simulated Data With A Graph

Exploring the simulated data with a graph helps us to ensure that we have simulated plausible data.

```r
ggplot(MICSsimulated,
       aes(x = cd1, # x is spanking
           y = aggression, # y is aggression
           color = factor(country))) + # color is country
  geom_smooth(method = "glm", # glm smoother
              method.args = list(family = "binomial"),
              alpha = .1) + # transparency for CI's
  labs(title = "Aggression as a Function of Spanking",
       x = "spank",
       y = "aggression") +
  scale_color_viridis_d(name = "Country") + # nice colors
  theme_minimal()
```



A.5 Explore The Simulated Data With A Logistic Regression

Similarly, exploring the data with a logistic regression confirms that we have created plausible data.

```
fit1 <- glmer(aggression ~ cd1 + cd2 + cd3 + cd4 +
                (1 | country),
              family = "binomial",
              data = MICSsimulated)

tab_model(fit1, # nice table
          transform = NULL) # untransformed estimates
```

aggression

Predictors

Log-Odds

CI

p

(Intercept)

0.76

0.54 – 0.98

<0.001

spank

0.19

0.03 – 0.36

0.019

beat

0.38

0.01 – 0.76

0.046

shout

0.42

0.26 – 0.58

<0.001

explain

-0.42

-0.62 – -0.23

<0.001

Random Effects

^{2}

3.29

$_{00\ country}$

0.04

ICC

0.01

N$_{country}$

30

Observations

3000

Marginal R^2 / Conditional R^2

0.025 / 0.037

A.6 Write data to various formats

Lastly, we write the data out to various formats: R, Stata, and SPSS.

```r
save(MICSsimulated,
     file = "./simulate-data/MICSsimulated.RData") # R

write_dta(MICSsimulated,
          "./simulate-data/MICSsimulated.dta") # Stata

write_sav(MICSsimulated,
          "./simulate-data/MICSsimulated.sav") # SPSS
```