On the Generalizability of Sex-Differences in Risk Attitudes*

Jan Feld
(Victoria University of Wellington)

Preliminary draft

May 23, 2024

Abstract

Sex-differences in risk attitudes are often treated as universal. I test whether this treatment is justified by analyzing 64 measures of risk attitudes which cover over 525,000 people in 120 countries. I analyze my rich dataset using meta-analysis methods. These methods allow me to correct for sampling error and estimate the percentage of countries in which men are more willing to take risks. My results suggest that sex-differences in risk attitudes are not universal but depend on the measure used. According to general self-assessment questions and lotteries in the gain domain, men appear to be more willing to take risks in nearly all countries. However, according to risk measures in the loss domain, there are no clear sex differences in risk attitudes. I also see larger sex-differences in rich and gender-equal countries. My results suggest that the academic literature leads to a distorted picture, because it disproportionately relies on measures and samples in which sex-differences in risk attitudes bigger and more consistent.

Keywords: Risk preferences, sex, gender, meta-analysis, meta-science

JEL classification: J16, D01, D81

* School of Economics and Finance, Victoria University of Wellington and IZA, jan.feld@vuw.ac.nz; I received valuable comments from Andrea La Nauze, Bob Reed, and seminar participants at the University of Otago, UCMeta at the University of Canterbury, Victoria University of Wellington (School of Psychology), the University of Queensland, and the 2023 AIMOS conference.

1. Introduction

The idea that men are more willing to take risks than women has been cited to describe a wide range of phenomena. For example, sex-differences in risk-attitudes have been proposed to explain men's greater acceptance of COVID-19 vaccines in five European countries, (Steinert et al. 2022), men's greater enthusiasm about autonomous vehicles in the United States (Nair and Bhat 2021), and men's more volatile voting behavior in Ghana (Weghorst and Lindberg 2013).

Many researchers appear to treat sex-differences in risk attitudes like sex-differences in height. Men are taller than women on average in the vast majority of contexts. For example, men are taller than women in all countries in the world (Max Roser, Cameron Appel, and Ritchie 2021). Men also appear to be taller than women in the vast majority of workplace, universities, and firms. [1] Could sex-differences in risk attitudes be similarly universal? The current state of the literature does not yet warrant this conclusion for two reasons.

First, estimated sex-differences in risk-attitudes differ a lot between studies. While many studies suggest that men are more willing to take risks, some show no statistically significant differences, and a few even suggest that women are more willing to take risks (Eckel and Grossman 2008; Croson and Gneezy 2009; Filippin and Crosetto 2016). This variety of results could be explained by differences in measures used between studies. Different measures are often only weakly correlated and risk attitudes appear to have general and domain specific components (Pedroni et al. 2017; Vieider et al. 2015; Dohmen et al. 2011; Frey et al. 2017). There is some evidence that sex-differences in risk attitudes also depend on the elicitation method. For example, Filippin and Crosetto (2016) show that measuring risk attitudes with the Holt-Laury method leads to smaller sex-differences than measuring risk attitudes with the investment game or the Eckel-Grossmann method (Eckel and Grossman 2002; Gneezy and Potters 1997; Holt and Laury 2002).

Second, the literature on sex-differences in risk attitudes relies disproportionately on samples of students from western, educated, industrialized, rich and democratic (WEIRD) countries (Henrich, Heine, and Norenzayan 2010). For example, among the 17 studies covered in the literature by Eckel and Grossmann (2008), X use student samples, X use

[1] Of course, men are not taller than women in all contexts. There are contexts in which women are taller by chance or due to sex-specific selection (like a university that is famous for producing good female basketball players and good male jockeys).

samples from WEIRD countries, leaving only X samples from non-WEIRD non-student samples. WEIRD populations, especially WEIRD students, have been shown to be particularly unrepresentative of humans more broadly (Henrich, Heine, and Norenzayan 2010).[2] While some studies have estimated sex-differences in risk attitudes in non-WEIRD and non-student samples (e.g. Vieider et al. 2015; Falk et al. 2018; Kachelmeier and Shehata 1992), we do not know yet if these samples show systematically different results.

The heterogeneity of findings in the literature is consistent with a range of different interpretations. At the one extreme, sex-differences in risk attitudes may be universal. The few studies that show women are more willing to take risks might just result from sampling error. In other words, with sufficiently large samples, we might find that men are more willing to take risks in all population and according to all measures. At the other extreme, sex-differences in risk attitudes may be highly context specific. With sufficiently large samples, we might find that *women* are more willing to take risks in many populations and for many measures. Even the idea that men are more willing to take risks on average may be an artefact of the particular measures and samples used in the literature. To learn how generalizable results are, we need to analyse estimates of sex-differences in risk attitudes across many different populations while explicitly accounting for sampling error.

I could conduct a meta-analysis to do exactly that. With standard meta-analysis methods, we could estimate the distribution of the unobserved "true effect" (i.e. average sex-differences in risk attitudes across contexts) while statistically accounting for sampling error. This estimated distribution might suggest, for example, that the men are more willing to take risks in all contexts and all estimates that show the opposite are a result of sampling error. I could also conduct moderator analyses to investigate in which contexts and for which measures sex-differences are most pronounced.

However, conducting a meta-analysis comes with its own set of problems. For example, if the academic literature predominantly relies on WEIRD student samples, those

[2] Henrich, Heine, and Norenzayan (2010) review how social science findings differ between different types populations (e.g. WEIRD vs other industrialized countries, industrialized vs small-scale societies). Perhaps unsurprisingly, they show some findings appear to be universal while others differ substantially between populations. What is more striking – at least to a WEIRD reader like me – is that even findings that "feel" universal are not. For example, Müller-Lyer illusion – a *visual* illusion about how context affects the perception of the length of a line – is virtually absent for San foragers. Furthermore, WEIRD populations are often outliers. For example, Müller-Lyer is strongest for US undergraduate students (compared to 14 small scale societies and European South African students). Overall, Henrich, Heine, and Norenzayan (2010) argue that WEIRD people, and in especially WEIRD students, are a particularly bad subpopulation for generalizing about the human species as a whole.

samples will disproportionately affect the average results. The estimated average effect might also be driven by publication bias if, for example, studies that find men are more willing to take risks are more likely to be published (Brodeur et al. 2023; Brodeur et al. 2016). Unfortunately, current statistical methods often struggle to correct for publication bias (de Gendre et al. 2023; Kvarven, Strømland, and Johannesson 2020). Estimating the variance of the unobserved effect is challenging because it is difficult to distinguish differences because of methodologies between studies with differences in true effects.[3] Furthermore, conducting a meaningful moderator analysis is challenging because measures and context tend to vary at the same time. In sum, meta-analyses provide useful statistical tools, but they struggle in practice because of bad input.

In this paper, I overcome these shortcomings with an approach that consists of two key elements. The first is to use meta-analysis methods with "quality-controlled" inputs. Instead of relying on estimates from the literature, I only use my own estimates from datasets that measure risk attitudes across different countries. My estimates are "quality-controlled" in three aspects: 1) they do not suffer from publication bias, 2) they are estimated with people from a wide-variety of countries, and 3) they use a consistent estimation method and the same measure of risk attitudes for all countries.[4] However, using a single measure of risk attitudes leaves my results vulnerable to be an artefact of the particular measure used. I avoid this concern with the second key element of my approach: triangulation, that is, the use of different methods to study the same phenomenon (Denzin 2015). I have access to several measures of risk attitude and all measures are used in many different populations. This feature of the data allows me to investigate according to which measures and in which populations we should expect men to be more willing to take risks.

I use seven publicly available datasets in my analysis. Each dataset contains at least one measure of risk attitudes which is available for people from many countries. Four datasets are large international surveys with one single measure of risk attitudes. Those are the Global Preference Survey (GPS; 80,000 people in 76 countries), the World Value Survey (158,000 people in 78 countries), the Gallup End of Year Survey (55,000 people in 65

[3] Recent "many-analysts" studies have shown that this is a valid concern. Researcher who are given the same data and same research question nevertheless often find substantially different results (Breznau et al. 2022; Huntington-Klein et al. 2021; Hoogeveen et al. 2023).

[4] The issue of "bad inputs" in meta-analysis is addressed in different ways by different researchers. Some restrict their meta-analysis to estimates they deem to be of higher quality (Jackson and Mackevicius 2023). Some use individual level data to be able to hold methods across different studies constant (Meager 2019; García and Heckman 2023). In medical sciences, those are referred to as individual patient meta-analysis (e.g. Seidler et al. 2023; Riley, Lambert, and Abo-Zaid 2010). Yet others, like this one, exclusively rely on their own data to be able to hold the methods constant and avoid publication bias (de Gendre et al. 2023).

countries), and the Survey of Health, Ageing and Retirement in Europe (136,000 people in 29 countries). Two additional datasets are from replication packages from studies measure risk attitudes with different measures across different countries. The study by Vieider et al. (2015) gives me access to 50 measures of risk attitudes (2,900 students in 29 countries). The International Test on Risk Attitudes by Rieger, Wang, and Hens (2015) gives me 8 measures of risk attitudes (4,900 students in 26 countries). The COVIDiSTRESS survey gives me access to two measures of risk attitudes (109,200 people in 41 countries). Overall, my datasets cover 525,000 people from 120 countries and 64 different measures of risk attitudes.

To investigate how generalizable sex-differences in risk attitudes are, I estimate their country-level distribution for each dataset-measure combination. Let me illustrate my approach with the GPS and its only measure of risk attitudes. First, I regress the measure of willingness to take risks on a male dummy variable for each country covered in the GPS. The resulting male coefficients shows average sex-gap in willingness to take risks in the sample for a given country. Those sample differences are estimates for the sex-gap in willingness to take risks in the populations of each country included in the GPS. The standard errors of the male coefficients quantify the uncertainty of these estimates. Second, I use the estimates and standard errors to estimate a random effects model. The random effects model assumes that the true unobserved country-level sex-differences willingness to take risks is normally distributed. I use restricted maximum likelihood to estimate the mean and standard deviation of this distribution. Third, I leverage the assumption that the unobserved "effect" is normally distributed to infer the share of countries in which men are more willing to take risks than women. For example, my estimates suggest that the distribution of country-level sex-differences in willingness to take risks in the GPS has a mean of 0.20 standard deviations (SD, positive values indicate men are more willing to take risks) and a standard deviation of 0.11 SD. These estimates imply that men are more willing to take risks in 97% of the countries covered in the GPS according to the only measure of risk attitudes included in the GPS.

My results reveal important heterogeneity in terms of measures and countries. The average sex-difference differs widely between measures, ranging from -0.14 SD to 0.33 SD. However, there are some clear patterns. Most types of measures either show men are more willing to take risks or no clear sex-differences. No type of measure suggests that women are more willing to take risks. Men appear near universally more willing to take risks according to general self-assessment questions ("In general, how willing are you to take risks") and lotteries in the gain domain. Risk measures using lotteries in the loss domain show no clear sex-differences. Furthermore, we see no meaningful differences between WEIRD and non-WEIRD countries. However, sex-differences in risk attitudes are generally more pronounced in rich, developed, and gender-equal countries.

My results explain why a causal reader of the literature might conclude that men are generally more willing to take risks. The empirical evidence on sex-differences in risk attitudes disproportionately relies on measures and samples (e.g. gain lotteries in rich countries) in which sex-differences in risk attitudes are most pronounced. Researchers reading several individual studies in this literature can easily conclude that sex-differences in risk attitudes are as universal as sex-differences in height. This is problematic, especially for researchers interested in risky situations involving losses and poor countries.

This paper relates to several literature reviews and multi-context studies that have grappled with the question of how generalizable sex-differences in risk attitudes are. I review those studies in depth in Section 2. More broadly, my paper relates to an emerging literature looking for ways to increase the reliability of scientific findings. Some of these efforts, like multi-lab replications, have focused on approaches that increase statistical power and reduce concerns about p-hacking and publication bias (Hagger et al. 2016; Klein et al. 2018). Others have used statistical methods to account for differences between studies like the quality of implementation (Angrist and Meager 2023; Angrist and Hull 2023). Yet others have put great effort into collecting diverse samples (Ruggeri et al. 2022; Vardy et al. 2022; Henrich et al. 2001).

I contribute to the literature in two ways. First, I provide the most thorough investigation of sex-differences in risk attitudes to date. My analysis of a uniquely rich data covering 64 measures of risk attitudes, from 120 countries. By comparison, all studies covered in the most cited literature reviews together cover only XX different countries and X measures. By leveraging meta-analysis methods, I explicitly account for sampling error and estimate variation in estimates that is due to differences in unobserved effects. By exclusively analysing estimates that I created, I do not have to worry about differences in methodology or publication bias. Second, I illustrate an approach to investigating generalizability of an effect. Applying meta-analysis methods to “quality controlled” inputs and triangulation allows us to make progress in other literatures as well.

2. Setting the stage

2.1. *Terminology and background*

I use the term “risk” to refer to the variance of potential outcomes and “risk attitudes” to refer to the degree to which people are willing to trade off the expected value of an outcome for a reduction in risk. The concept of risk attitudes is conveniently illustrated with the choice between a sure amount of $5 and 50% chance of winning $10 and a 50% chance of winning nothing. Both options have the same expected value of $5, but the second option is riskier. People who prefer the safe option are “risk-averse”, people who are indifferent are “risk-neutral”, and people who prefer the risky gamble are “risk-seeking”. People are “more willing to take risks” if they need to be compensated (in terms of expected value) for taking on risks. People also face non-financial choices in which they trade off their expected values and risks. For example, people may be trading-off expected travel time and risk when deciding which route to take to work. Or, they may be trading off expected healthy life years and risk when choosing between different medical treatments. The concept of risk attitudes covers those contexts as well.

The two most prominent theories that aim to explain people’s risky choices are expected utility theory and prospect theory (Kahneman and Tversky 1979). Expected utility theory predicts that people are risk averse because of the decreasing marginal utility of wealth. For this reason, the curvature of the utility function is sometimes used synonymously with risk aversion. Because expected utility theory has been a poor predictor of people’s risky choices, Kahneman and Tversky (1979) have proposed prospect theory as an alternative. Prospect theory has two components: a value function and a probability weighting function. The value function describes how objective value translates into subjective value relative to a reference point. The probability weighting function describes how people translate objective probability into probability weights. People’s risky choices can be explained by the shape of the value function and the shape of the probability weighting function. For example, the convex shape of the value function in the loss domain has been used to explain why people are risk seeking for losses. And overweighting of small probabilities has been used to explain why people find lotteries attractive.

There are several reasons men may be more willing to take risks than women. Sex-differences in risk attitudes can result from differences in the curvature of the utility function, the shape of the value function, or the shape of the probability weighting function.
At a deeper level, sex-differences in risk attitudes may be driven by different evolutionary pressures for women and men (Dekel and Scotchmer 1999) and differences in testosterone levels (Sapienza, Zingales, and Maestripieri 2009). They can also reflect social norms and gender roles around risky behavior.

When considering the wide variety of explanations, it is also plausible that sex-differences in risk attitudes may differ between measures and populations. For example, there may be sex-differences the shape of the value function may differ in the gain domain but not in the loss domain. This would predict that sex-differences in risk attitudes would only appear in lotteries in the gain domain. Or differences in social norms and gender roles can explain why sex-differences in risk attitudes differ between countries.

2.2. Literature review

Many studies have investigated sex-differences in risk attitudes.[5] The general approach of these studies is straightforward: they measure risk attitudes and see if women and men score on average differently on that measure. What do these studies tell us about sex-differences in risk attitudes?

Three literature reviews have attempted to answer this question. All suggest that men are more willing to take risks on average. However, these three imply different conclusions about how ubiquitous sex-differences in risk attitudes are. The review by Croson and Gneezy (2009) suggests that men are near universally more willing to take risks. The authors summarize 13 estimated sex-differences in risk attitudes from 10 studies. Of those 13 estimates, 11 show men are more willing to take risks and 2 show no sex-differences. After reviewing this evidence, the authors conclude that men are more willing to take risks in "*the vast majority of environments and tasks*" (Croson and Gneezy 2009, page 449). Less conclusive is the literature review by Eckel and Grossman (2008). The authors summarize 28 estimated sex-differences in willingness to take risks from 17 experimental studies. Of these 28 estimates, 14 show men are significantly more willing to take risks, 2 show women are more willing to take risks, and the remaining 12 estimates are not even statistically significant at the 10% level. In their Table 1, the authors break down how effects differ depending on the domain (gain or loss), the frame of the experiment (abstract or contextual) and type of study (lab vs field). They conclude that most studies find men are more willing to take risks than women, but studies with contextual frames show less consistent results. Niederle (2017)

[5] The literature on sex-differences about willingness to take risks uses a variety of different terms to refer to the same basic phenomenon. For example, some studies refer to "risk attitudes" as "risk aversion" or "risk preferences"; some studies refer to "sex" whereas others refer to "gender". All studies have in common that they describe differences in measures of willingness to take risks between people who self-identify (or are judged by the interviewer as) female or male. For consistency reasons, I will use the term "sex" and "risk attitudes" when describing studies which use different terminology. For example, while Falk et al. (2018) report "gender differences in risk preferences", I will describe their results as showing "sex-differences in risk attitudes".

reflects on the other two literature reviews and summarizes the results of two multi-context studies not covered in the other reviews. She suggests that the large heterogeneity in results may reflect small sex-differences in risk attitudes in combination with small samples as well as differences due to elicitation methods. Her conclusion is most careful. She suggests researchers should not just assume sex-differences in risk attitudes by measuring it for the task at hand. [6]

It is hard to assess the generalizability of sex-differences in risk attitudes with literature reviews for two reasons. First, the studies included in the literature are likely unrepresentative of all possible studies. Researchers decide in which context to conduct their study and how to measure risk attitudes. Even among all conducted studies, some do not get written up, some never get published, and some published studies may get overlooked in literature reviews. Non-random selection at any of these stages can result in a distorted picture. For example, of the 28 estimated sex-differences in risk attitudes reported on by Eckel and Grossman (2008), X were based on WEIRD samples, X on students, and only X on non-WEIRD non-student samples. This reliance on WEIRD samples may be problematic, because WEIRD populations have been shown to be particularly unrepresentative for humans more generally. The results may not even be representative of the mostly WEIRD populations covered in the literature review if there is publication bias. For example, if studies finding men are more willing to take risks are more likely to get published (or included in the literature review) we may overestimate the degree to which is the case even for WEIRD student populations.

Second, it is hard to discern which study characteristic matters for sex-differences in risk attitudes because individual studies differ on many dimensions (e.g. population, measure of risk attitudes). This problem is aggravated by the relatively small number of studies summarized in the literature reviews. For example, among the studies reviewed by Eckel and Grossman (2008) all three field studies show that men are more willing to take risks. However, is impossible to know if these results are the field nature of the study, others study characteristics (e.g. population), or chance.

[6] Byrnes, Miller, and Schafer (1999) conduct a meta-analysis to summarize 150 studies on sex-differences in risk-taking (not risk attitudes). They define risk taking to involve "*the implementation of options that could lead to negative consequences*" and summarize studies that measure risk taking with a wide range of different behaviours such as drug-taking and driving behavior. The results of this meta-analysis suggest that men are more risk-taking but results differ markedly by context. Note that risk-taking differs from risk attitudes. Risk attitudes are not limited for risk situations that involve negative consequences. For example, in the Holt and Laury (2002) method participants can choose between different lotteries in the gain domain and do not face the risk of losing money.

A complementary approach for learning about the generalizability is therefore studies that use the same measure of risk attitudes across many contexts. There are five such multi-context studies. Two of them suggest men are near universally more willing to take risks. Charness and Gneezy (2012) analysis sex-differences in willingness to take risks of 15 studies which use variations of the same investment game to elicit risk attitudes. They show that men are more willing to take risks in 14 of these studies. The samples in these summarized studies are relatively diverse. They come from 8 countries (US, Sweden, Germany, Netherlands, China, India, Turkey, Tanzania) and also include non-student samples (e.g. tournament bridge players, villagers in China). Falk et al. (2018) measure willingness to take risks with a composite measure comprising a hypothetical lottery choice and one self-assessment question across representative samples adults in 76 countries. The countries varied widely in terms of geography, culture, and economic development. In 72 of these countries, the point estimates suggest men are more willing to take risks. In 59 of those, the estimated sex-difference is statistically significant at the at least at the 10% level.

In contrast, Filippin and Crosetto (2016) find more heterogenous results. The authors gathered data from 63 published and unpublished studies using the Holt and Laury (2002) measure of risk attitudes. While Filippin and Crosetto (2016) do not break down the samples of the included studies, my spot checks suggest that most of them were university students. Of the 63 studies, only 8 (12.6%) showed that men are significantly more willing to take risks. The remaining studies showed no statistically significant sex-differences. Furthermore, among the 54 studies for which individual-level data is available, the pooled estimates suggest that men are only 0.17 SD more willing to take risks—as an estimate that is much small than findings reported in other studies. Filippin and Crosetto (2016) conclude that, when using the Holt-Laury measure, sex-differences in risk attitudes are not a ubiquitous phenomenon.

Two studies use student samples from large and diverse countries. While not at the focus of their studies, these also estimate sex-differences in risk attitudes for their pooled sample. Rieger, Wang, and Hens (2015) use data from 53 countries and find that men are, on average, more willing to take risks for gains and less willing to take risks for losses. Vieider et al. (2015) estimate sex-difference in risk attitudes for 11 different measures and find men are more willing to take risks according to 7 of them with no significant sex-differences according to the other 4 measures.

I have four take-aways from these literature reviews and multi-context studies. First, sex-differences in risk attitudes should not be treated as universal. The existing evidence points to meaningful differences depending on the measures used: some measures suggest men are near universally more willing to take risks, others suggest show much more mixed results, with some even suggesting that women are on average more willing to take risks. Second, we do not know yet to what extent sex-differences differ between different populations (e.g. WEIRD vs non-WEIRD populations). While several studies have included a wide people from very diverse sets of countries, none has explored if results differ for different types of samples. Third, part of the variation in estimates reflects sampling error. This problem is likely particularly pronounced in experimental studies which use relatively small samples. With small samples, one should expect that some studies find insignificant results and some even find results with the wrong sign. Fourth, estimates are difficult to compare between studies because of subtle differences in methodologies. For example, Filippin and Crosetto (2016) show estimated sex-differences without controls, Rieger, Wang, and Hens (2015) include controls for log of GDP per capita, Vieider et al. (2015) include 10 different control variables, and Falk et al. (2018) control for age, age-squared, and subjective math skills. Several many-analysts project has shown that seemingly innocuous methodological choices (e.g. about control variables and sample restrictions) can have meaningful effects on estimates (Breznau et al. 2022; Huntington-Klein et al. 2021; Hoogeveen et al. 2023).

To get a more conclusive answer whether and in which circumstance men are more willing to take risks, I will do the following. First, I will analyse sex-differences in risk attitudes with richer data than any previous study. My data consists of 7 datasets: three datasets from the multi-context studies by Rieger, Wang, and Hens (2015), Falk et al. (2018), and Vieider et al. (2015), as well as additional data from 4 datasets. Second, I will use the same methodology across different datasets to increase comparability. Third, I will also delve deeper into sex-differences than previous studies that have analysed the same data. For example, Vieider et al. (2015) present 11 estimated sex-differences in risk attitudes for their pooled dataset. I will estimate sex-differences in risk attitudes for all measure-country combinations (1,980 estimates in total). Fourth, I will explicitly correct for sampling error using meta-analysis methods. Fifth, I will explore to what extent sex-differences in risk attitudes differ between measures and populations.

3. Dataset and Risk measures

3.1. Datasets

Table 1 shows the summary statistics for the seven data included in my analysis. Across all 7 datasets, I have access to 64 different measures of risk attitudes from 525,886 people in 120 countries. Each dataset has its own strengths and weaknesses. For example, some give me access to large and representative samples of many countries, but only one measure of risk attitudes. Other datasets only give me access to many measures of risk attitudes, but only use student samples.

Table 1: Summary statistics

(1) Data source	(2) Number of countries	(3) Number of people	(4) Number of measures	(5) Mean age	(6) Share male	(7) Sample
GPS	76	80,337	1	41.82	0.45	representative
WVS	78	161,293	1	41.77	0.48	representative
Gallup	65	55,533	1	41.54	0.50	representative
SHARE	29	111,704	1	65.15	0.44	representative 50+
VIE	29	2,874	50	21.83	0.52	student
INTRA	26	4,945	8	21.26	0.53	student
Covid	41	109,200	2	38.70	0.27	convenience
combined	120	525,886	64	40.96	0.45	

Figures 1 and 2 show the countries covered for each dataset. Each of the first three datasets contain large representative samples from a diverse set of countries. The Global Preference Survey (GPS) gives me access to representative samples from 80,337 people in 76 countries (target population: 15+ years), the World Value Survey (WVS) gives me access to a representative sample of 161,293 people from 78 countries (target population: between 18 and 85 years), and the Gallup End of Year Survey (hereafter Gallup survey) gives me a representative sample of 55,533 people from 65 countries (target population: 15+ years). In each of the three datasets, the mean age is around 42 years and the men make up between 45 and 50 percent of the sample. All three datasets contain one risk measure and were collected via face-to-face or phone interviews.

Figure 1: Countries included in any dataset, GPS, WVS, and Gallup datasets

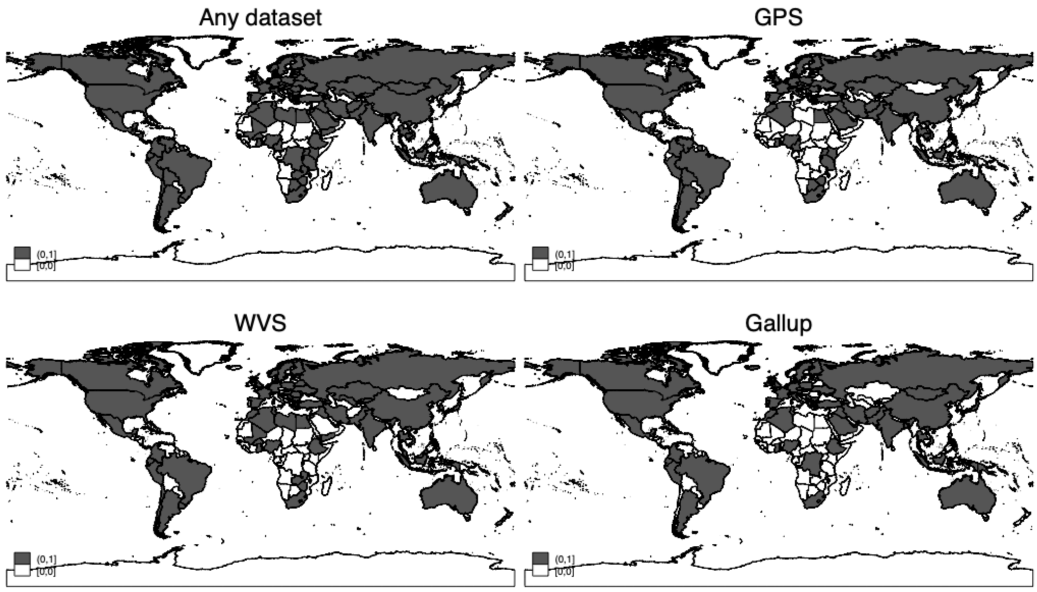
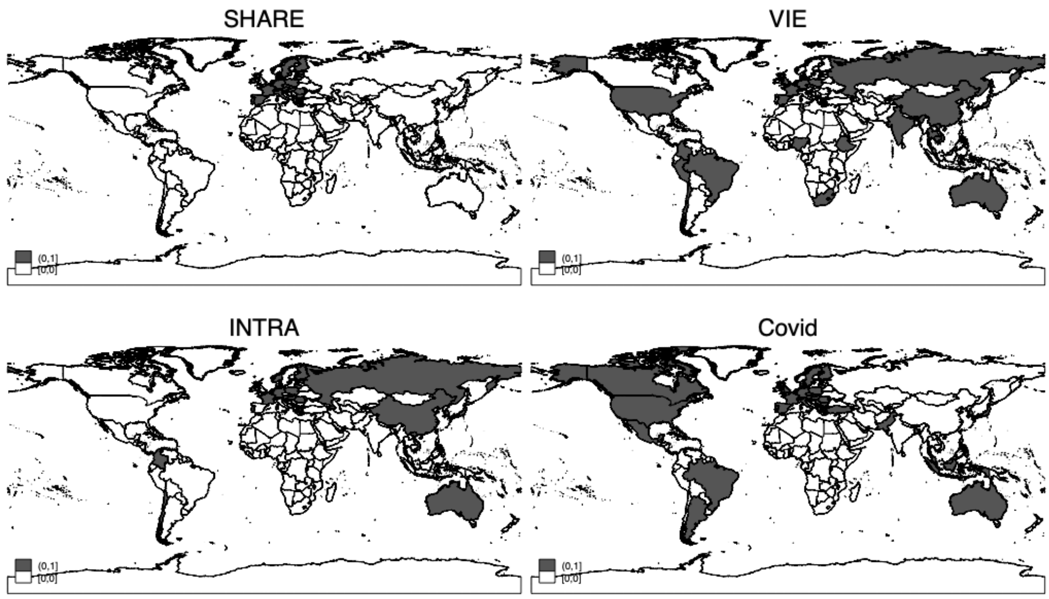


Figure 2: Countries included in SHARE, VIE, INTRA, and Covid datasets



The data from the Survey of Health, Ageing, and Retirement in Europe (SHARE) also includes one measure of risk attitudes (Börsch-Supan et al. 2013). However, SHARE is limited to European countries and aims to be representative of all people in a given country who are 50 years or older. The SHARE data is collected via in-person interviews. It covers 111,704 people from 29 countries. The mean age is 64 years and men make up 44 percent of the sample.

Two additional data sources come from publicly available replication packages of studies which have measured risk attitudes in student samples from many countries. The study Vieider et al. (2015) measures risk attitudes in 50 different ways. After excluding data from Saudia Arabia which only contained male students, the Vieider et al. data (VIE data) covers of 2,874 students from universities in 29 countries. Data of the International test of risk attitudes (INTRA) comes from the replication package of Rieger, Wang, and Hens (2015). The replication package contains data from 60 universities in 53 countries. I limit my sample to countries with data from at least 100 students for each of the 8 measures of risk attitudes. These sample restrictions leave me with 4,945 students from universities in 26 countries. The average age in the VIE and INTRA data is around 21 years and the share of men is around 52 percent.

The final dataset comes from the COVIDiSTRESS survey (hereafter Covid survey) which was conducted between March 30th and May 30th of 2020 during the early stages of the Covid pandemic (Yamada et al. 2021). The Covid survey was organized by a global consortium of researchers from 39 countries, respondents were recruited via online and media appeals. It contains two measures of risk attitudes. After limiting my sample to countries for which each risk measure is available for at least 100 respondents, the Covid survey gives me access to data from 109,200 people from 41 countries.

3.2. Measures

My 64 measures of risk attitudes range widely in the domains they cover and how they were administrated (see Table 2 for an overview).

Table 2: Summary of Risk Attitudes Measure

(1) Data source	(2) Short description	(3) Details
GPS	1 composite measure	composite of one self-assessment general-risk question and one lottery
WVS	1 self-assessment question	Importance of adventure and risk taking in one's person's life
Gallup	1 risk lottery in the gain domain	Choice between 50/50 chance of doubling household income or 50% increase for sure
SHARE	1 self-assessment question	willingness to trade off risk and return when saving or investing
VIE	13 risk lotteries in the gain domain	range of potential outcomes: €0 euros to €20 range of probabilities: 0.125 to 0.875 range of expected values: €2.5 to €18.13
	8 uncertainty lotteries in the gain domain	range of potential outcomes: €0 euros to €20 range of suggested probabilities: 0.125 to 0.875 range of suggested expected values: €2.5 to €18.13

	13 risk lotteries in the loss domain	range of potential outcomes €0 euros to €20 euros range of probabilities: 0.125 and 0.875 range of expected values: €-2.5 and €-18.13
	8 uncertainty lotteries in the loss domain	range of potential outcomes €0 euros to €-20 euros range of suggested probabilities: 0.125 and 0.875 range of suggested expected values: €-2.5 and €-18.13
	8 self-assessment questions	answer scales range from 0 "not at all willing to take risks" to 10 "very willing to take risks" domains: in general, driving, financial, sport, occupation, health, people
INTRA	6 risk lotteries in the gain domain	range of potential outcomes: €0 euros to €10,000 range of probabilities: 0.1 to 0.9 range of expected values: €10 to €60,000
	2 risk lotteries in the loss domain	range of potential outcomes: €0 euros to €-100 range of probabilities: 0.4 to 0.6 expected values: €-48 and €-60
Covid	1 risk lottery in the gain domain	Support for saving 200 lives for sure (policy A) or a 1/3 chance of saving 600 lives (policy B)
	1 risk lottery in the loss domain	Support for saving 400 deaths for sure (policy C) or a 2/3 chance of 600 deaths (policy D)

The GPS risk measure was developed with an extensive validation procedure by Falk et al. (2023). The goal of the procedure was to find a risk measures that are easy to implement in large scale surveys and are good predictors of incentivized measures typically used in laboratory experiments. The GPS risk measure is based on two items. The first item is a hypothetical lottery in which respondents are given five binary choices between varying safe amounts and a lottery. The lottery is the same across all five choices, and the save amount varies depending on the respondent's previous answer in a way that zooms in—via the "staircase method"–to each respondent's certainty equivalent of the lottery. The amounts differed between countries according to the median household income in a country. For example, in Germany the lottery consisted of a 50% chance of winning 300 euros and a 50% chance of winning nothing, and the first binary choice was between a sure amount of 160 euros and the lottery. The second item is the question "In general, how willing you are to take risks?", which respondents could answer on an 11-point scale ranging from 0 ("completely unwilling to take risks") to 10 ("very willing to take risks"). The final risk measures is the weighted average of the two items, with the weights stemming from the validation procedure.

The WVS risk measure consists of one item which is elicited as follows: As part of a 10 item version of the Schwartz Value Inventory (Schwartz 2012), the interviewer describes a gender-matched person as "Adventure and taking risks are important to this person; to have an exciting life." The interviewee is asked to indicate how much this person is like them by showing one of six different cards. The answer possibilities on the cards range from "very much like me" to "not at all like me". Bouchouicha and Vieider (2019) have shown that the WVS risk measure is correlated in the expected direction with other risk measures in two ways. First, they have shown that the WVS measure is correlated at the individual level with certainty equivalents from incentivized lotteries and the answers to the general-willingness-to-take-risks question in an Indian sample. Second, country-level averages of the WVS risk measure correlate with country-level averages of six different risk measures taken from Vieider et al. (2015).

The Gallup risk measure consists of the following question: *"Think about your current household income: which of the following choices would you choose if offered?* Respondents could choose between the following two answers: *"A 50/50 chance to receive double your household income"* and *"A guaranteed increase in your household income of 50 %"*.

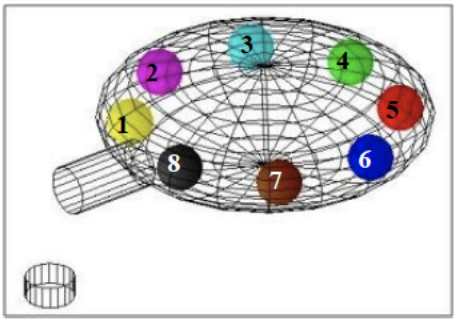
The SHARE risk measure is about how much risks people are willing to take in their saving and investment decisions. Respondents are reminded about the risk and reward trade-off between different investment options and are then asked which of the following four statements best describes their risk attitude:

- "*Take substantial financial risks expecting to earn substantial returns",*
- *"Take above average financial risks expecting to earn above average returns",*
- *"Take average financial risks expecting to earn average returns",* and
- *"Not willing to take any financial risks"*

The study by Vieider et al. (2015) gives us access to 50 measures of risk attitudes which were elicited in-person with pen-and-paper questionnaires. Of those 50 measures, 43 come from incentivized lotteries. In all lotteries, students were shown one lottery with two potential outcomes and their certainty equivalent for this lottery was then measured with a series of binary questions about whether they preferred a fixed amount or the lottery. As an example, Figure 3 shows the first lottery measure in which students had a 50% chance of winning €5 and a 50% chance of winning nothing. The probabilities were described as chances of drawing a ball with the numbers 1,2,3, or 4 from an urn that has 8 balls numbered 1-8. The first binary question asked whether students prefer the lottery or €0.50 for sure, and the sure amount increased in subsequent questions until €4.50. The estimate of students' certainty equivalent is the average of the sure amounts around students' switching points. For example, if students preferred the lottery over €2.00 for sure, and €2.50 for sure over the lottery, the estimated certainty equivalent is €2.25.

Figure 3: Example of VIE lottery

Decision 1



Lottery	Sure	
O	O	€ 0.50 for sure
O	O	€ 1.00 for sure
O	O	€ 1.50 for sure
O	O	€ 2.00 for sure
O	O	€ 2.50 for sure
O	O	€ 3.00 for sure
O	O	€ 3.50 for sure
O	O	€ 4.00 for sure
O	O	€ 4.50 for sure

Win **€ 5** if one of the following balls is extracted:

1 2 3 4

Win **€ 0** if one of the following balls is extracted:

5 6 7 8

The VIE lotteries belong to four distinct categories. The first category is “risk lotteries in the gain domain” in which students could win money and the probabilities were clearly shown (13 lotteries). The second category is “uncertainty lotteries in the gain domain” in which students could win money, but the probabilities of winning were unclear (8 lotteries). As for the risk lotteries, those lotteries described an urn with 8 balls and showed which numbers mean winning the lottery (e.g. win €5 if balls 1,2,3 or 4 are drawn). However, the urn shown on the questionnaire was opaque. Students were told that this urn also had 8 balls but they cannot know the numbers in this urn (“*All balls bear a number between 1 and 8 inclusive (have either 1 , 2 , 3 , 4 , 5 , 6 , 7 , or 8 written on them), but it is possible that some numbers are absent from this urn while others occur repeatedly.*”). In Table 2, I describe the “suggested probabilities” of those lotteries assuming that the urn balls numbered 1-8. The third category is “risk lotteries in the loss domain” in which students could lose up to €20 (13 lotteries). Students were endowed with €20, so they risked losing “house money”. The fourth category is “uncertainty lotteries in the loss domain” in which students could lose money and the probabilities were suggested.

The VIE dataset also includes 7 self-assessment risk measure which asked to students about their willingness to take risks “in general” and in seven specific domains (driving, financial, sport, occupation, health, people). The answer scale for all 8 questions ranged from 0 "not at all willing to take risks" to 10 "very willing to take risks".

The INTRA dataset contains eight lottery measures of risk attitudes, six in the gain domain and two in the loss domain. Each lottery had two potential outcomes and students were asked to indicate the maximum amount they are willing to pay for the lottery. All questions were hypothetical—there was no money on the line.

The CS dataset consists of two measures of risk attitudes, one in the gain domain and one in the loss domain. Both measures come from a modified version of the "Asian Disease Problem" (Tversky and Kahneman 1981). Respondents were randomly assigned one of two scenarios in which had to state their policy preference for a hypothetical outbreak of an unusual disease.[7] Respondents in the gain frame were shown the following two options:

- *"If Program A is adopted, 200 people will be saved.*
- *If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved"*

Respondents in the loss frame were shown these two options:

- *"If Program C is adopted 400 people will die.*
- *If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die."*

Note that the scenarios are numerically identical. All four programs have the same expected number of saved lives (200) and the second option shown (Programs B and D) are riskier.

4. Empirical strategy

My empirical strategy consists of two parts. In the first part, I estimate sex-differences in risk attitudes and their standard errors for all country-measure combinations. In the second part, I use those estimates and standard errors to estimate the distribution of sex-differences in risk attitudes for each measure.

Estimating sex-differences in risk attitudes for country-measure combination

[7] The exact question wording is "*"Imagine that [your country] is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows: Which of the two programs would you favour"*

For each country-measure combination, I estimate the following model

$$risktaking_{icm} = \alpha_{cm} + \beta_{cm} male_i + u_{icm} \quad (1)$$

where $risktaking_{icm}$ is the score of person i in country c according to measure m. This measure is standardized to have a mean of zero and standard deviation of one among all people in my sample for who this measure is available. $male_i$ is a dummy variable, which is 1 if the person is male and 0 if the person is female. α_{cm} represents the average value of risk-taking of women in the population in country c according to measure m. The parameter of interest is β_{cm} which represents the average sex-difference in the population in country c according to measure m. Positive values of β_{cm} indicate that men are more willing to take risks. The error term u_{icm} represents the deviation of person i from the sex-specific population average risk-taking score in country c according to measure m.

I estimate Equation (1) with ordinary least squares (OLS) regressions with heteroskedasticity robust standard errors. The OLS estimate $\hat{\beta}_{cm}$ is the sample average sex-difference in risk attitudes in country c for measure m, and the standard error $se(\hat{\beta}_{cm})$ is a measure of the uncertainty of this measure.[8]

A key assumption is that error term u_{icm} is unrelated to a person's sex ($E[u_{icm}|male_i] = E[u_{icm}]$). In other words, I assume that sex-differences in risk attitudes in the sample are not systematically different from the sex differences in risk attitudes in the population. This is a plausible assumption which underlies, often implicitly, all empirical studies on sex-differences in risk attitudes. There might be a systematic selection of people into my sample. Such a selection would only be a problem if it is related to sex and risk taking. For example, my estimates would be biased if risk-seeking women *-but not men-* avoid taking part in household surveys or experiments. While such selection is conceivable, I do not believe it is empirically important.

Estimating the unobserved distribution of population-level sex-differences in risk attitudes

[8] Note that for the SHARE dataset I observe multiple observations per person. I keep all observations in my analysis and cluster my standard error at the individual level.

I follow the meta-analysis literature and describe the distribution of each estimate for each measure with a random effects model as follows:

$$\hat{\beta}_{cm} = \beta_{0m} + \theta_{cm} + \epsilon_{cm}, \qquad (2)$$

where $\hat{\beta}_{cm}$ is the estimated sex-differences in risk attitudes for country *c* according to measure m, β_{0m} is the mean sex-differences in risk attitudes according to measure *m* across all countries for which this measure is available (the "grand mean"), θ_{cm} indicates country specific deviations from the grand mean, and ϵ_{cm} indicates sampling error, that is the differences between a country's true sex-differences according to measure *m* ($\beta_{0m} + \theta_{cm}$) and the observed estimate ($\hat{\beta}_{cm}$).

I further assume that the true country-level distribution of sex-differences in risk attitudes according to measure m is normally distributed with a mean of β_{0m} and a variance of τ_m^2, and the sampling ϵ_{cm} is normally distributed with a mean of zero and a variance of $se(\hat{\beta}_{cm})^2$.

$$\beta_{cm} \sim N(\beta_{0m}, \tau_m^2) \qquad (3)$$

$$\epsilon_{cm} \sim N\big(0, se(\hat{\beta}_{cm})^2\big) \qquad (4)$$

Note that the distribution shown in Equation (3) is the target of my empirical analysis. Instead of estimating the sex-difference in risk attitudes in a single country, my goal is to estimate the distribution of population level sex-differences between countries.

I estimate Equation (3) with restricted maximum likelihood, using the coefficients and standard errors for each country as input. I implement this estimation using the *meta summarize* command in Stata 18.

5. Results

5.1. Distributions of "raw" sex-differences in risk attitudes

After estimating Equation (1) for all country-measure combinations, I have 1,980 coefficients of interest. Each of those represents a sample difference in risk attitudes for one measure in one country. Figures 4 and 5 show the means distributions of these coefficients jointly, as well as separately for each dataset. These figures reveal three things. First, the mean of all coefficients is positive in each dataset, suggesting that men are on average more willing to take risks. Second, there is substantial variation in the coefficient estimates. The standard deviation of all coefficients is 0.22 SD, which is larger than the overall mean of 0.10 SD. Third, there is substantial heterogeneity between datasets. For example, the means range from 0.03 SD (Gallup and Covid) to 0.023 SD (WVS and SHARE). In two datasets (WVS and SHARE) all estimates are positive, whereas in four datasets (Gallup, VIE, INTRA, and Covid) more than 30% of coefficients are negative.

Figure 4: Coefficients of interest from all datasets combined, GPS, WVS, and Gallup

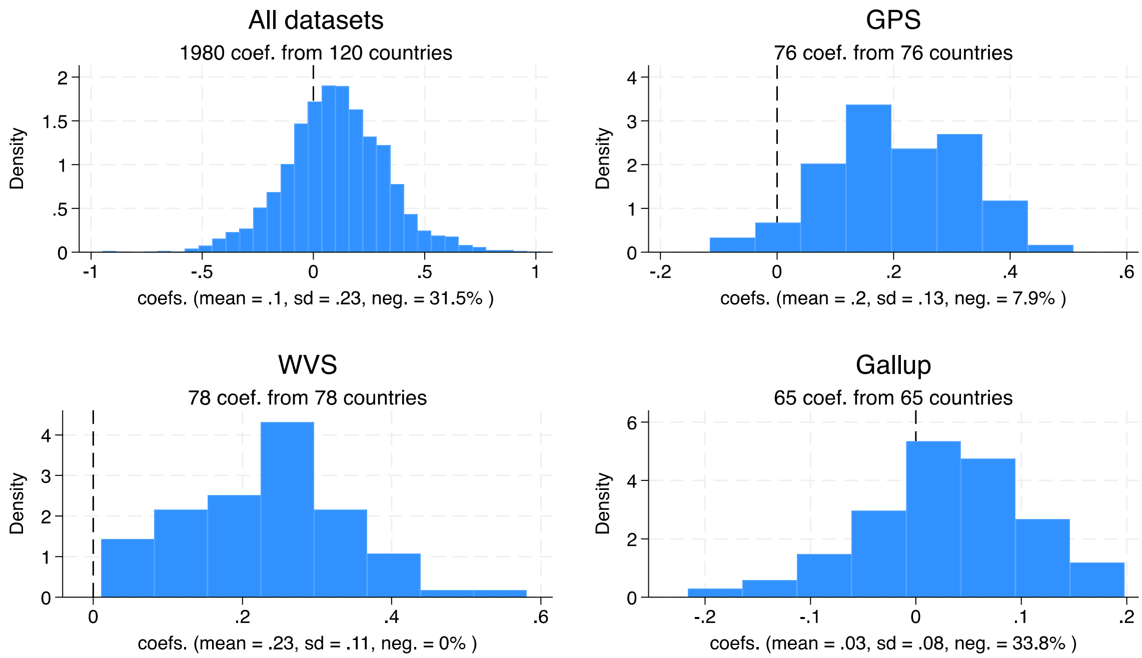
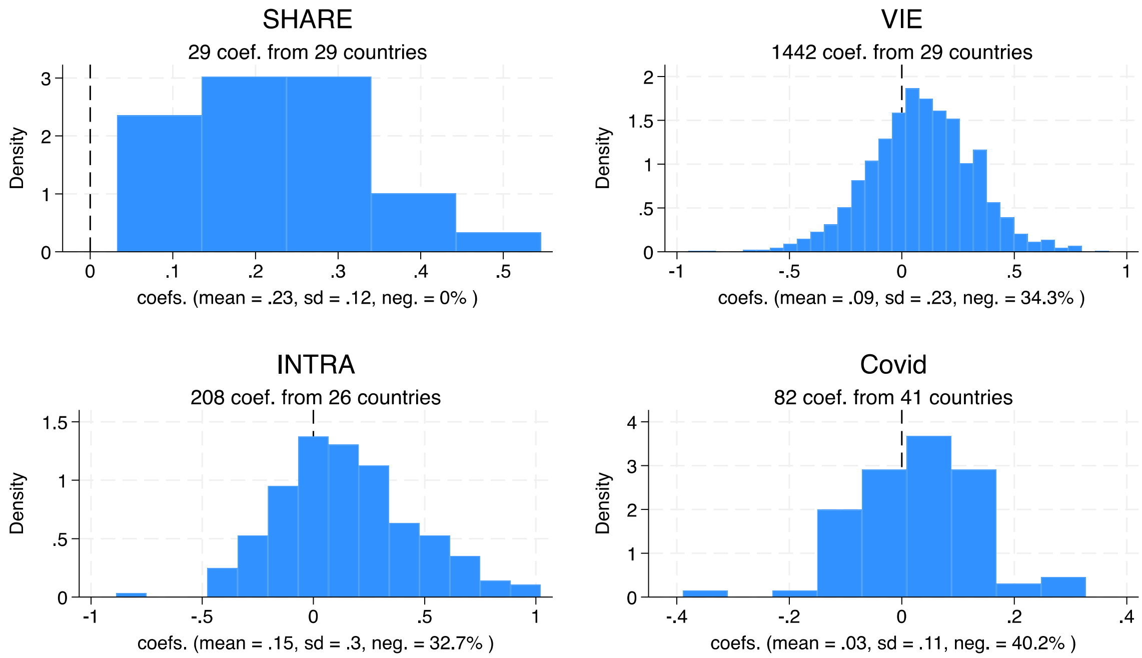


Figure 5: Coefficients of interest from SHARE, VIE, INTRA, and Covid



The differences in means between datasets suggest that sex-differences in risk attitudes differ between measures or populations. The differences in standard deviations may reflect differences in true heterogeneity as well as reflect sampling error. Some of the coefficients were estimated with only 100 people, while others were estimated with several thousand people. It is therefore unsurprising that the two datasets with the smallest sample sizes (VIE and INTRA) have the largest standard deviation of coefficients.

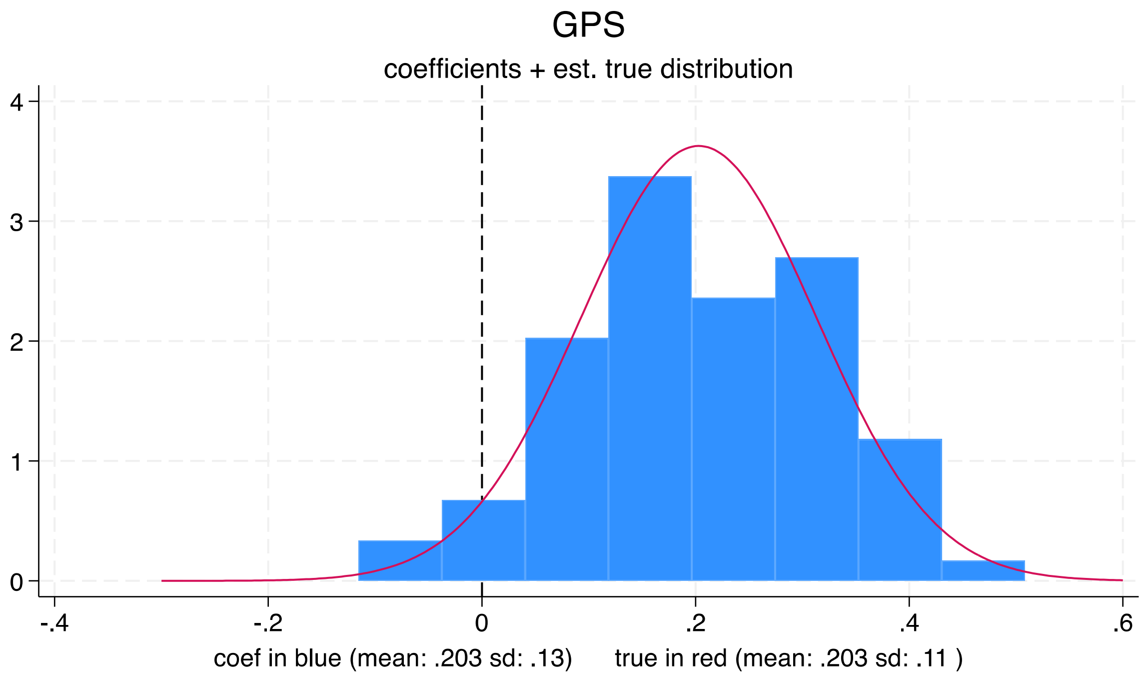
5.2. Random effects model estimates: general approach

A standard approach in meta-analyses is to estimate an overall mean "effect" and standard deviation of the effect using all estimates. As a baseline, I follow this convention and estimate a random effects model summarizing all 1,980 coefficients of interest. The results suggest men are on average 0.11 SD more willing to take risks (p-value <0.001). The estimated standard variance of the "true-effect size", that is, the sex-differences in risk attitudes across all estimates after accounting for sampling error is 0.02 SD. Leveraging the normality assumption, this mean and standard deviation imply that 78% of all true effects (each of which reflect a country-measure combination) is positive. This percentage is larger than the 68% of estimates that are positive, reflecting the correction for sampling error.

However, estimates of this exercise are hard to interpret because they combine different measures and different populations (see Data colada posts). Furthermore, data from the experimental studies are overrepresented because they have contributed many measures per participant. A cleaner approach is to estimate sex-differences in risk attitudes for each individual measure. This approach has the advantage that all measures are comparable and the sample is clearly defined.

Let me illustrate this approach with GPS data for which only one measure of risk attitudes is available (see Figure 6). My random effects model estimates suggest that the true country-level sex-differences in risk attitudes have a mean of 0.20 SD and a standard deviation of 0.11 SD, which implies that men are more willing to take risks in 97% of countries. These results are much easier to interpret. The mean and standard deviation come from a single measure of risk attitudes and the sample is clearly defined and transparent (representative sample of 76 countries, shown in Figure 1).

Figure 6: Illustration of empirical approach with GPS data



Restricting my analysis to GPS data has the disadvantage that my conclusions hinge on the particular measure used. I therefore repeat this exercise with for each available measure and save the resulting estimated true mean, true standard deviation, and percent of countries in which men are more willing to take risks. Instead of performing one meta-analysis with 64 different measures, I perform 64 meta-analyses holding the measure constant for each one of those (but allowing it to differ between them).

5.3. Means, standard deviations, and share positive for all 64 measures

Figure 7 shows the means and standard deviations of the estimated distributions of country-level sex-differences in risk attitudes for all 64 measures. The majority of the means (83%) are positive. Put differently, men are more willing to take risks than women on average, according to most measures. The standard deviations vary substantially between measures. The smallest estimated standard deviation is 0.0001 SD for an uncertainty lottery in the gain domain from the VIE dataset. For this particular risk measure, my estimates suggest the population level-sex differences in risk attitudes are almost identical among all 29 countries in the VIE dataset. By contrast, the largest estimated standard deviation is 0.2926 SD for a risk lottery in the gain domain from the INTRA dataset. For this measure, there appears to be substantial variation in sex-differences in risk attitudes among 26 countries included in the INTRA dataset.

Figure 7: Estimated means and standard deviation, by measure

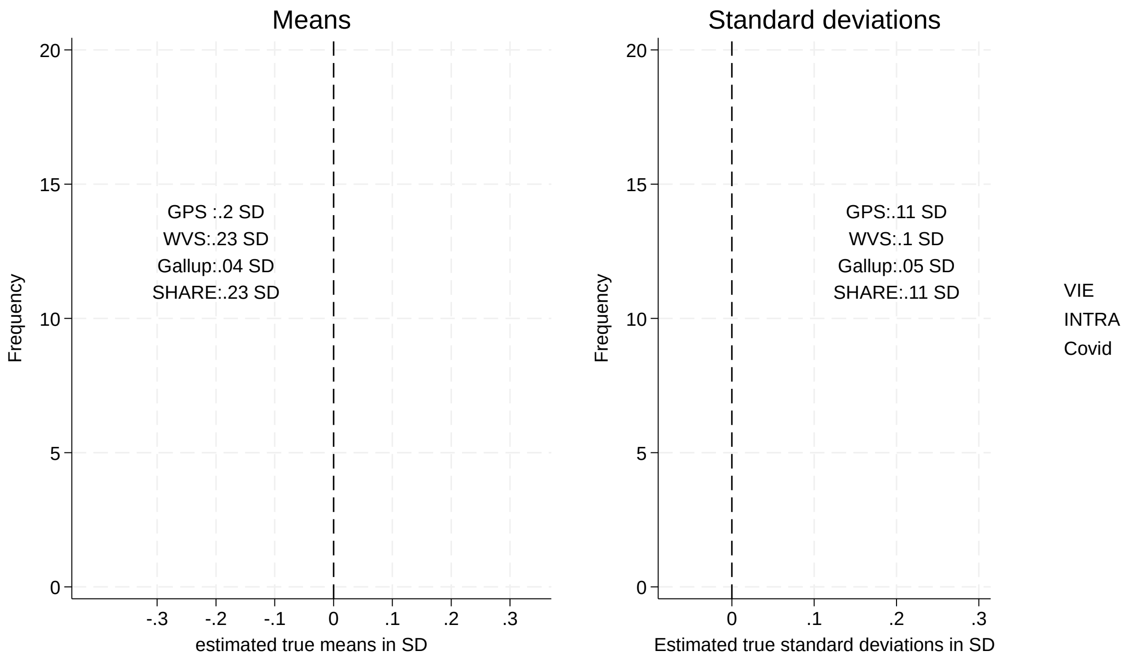
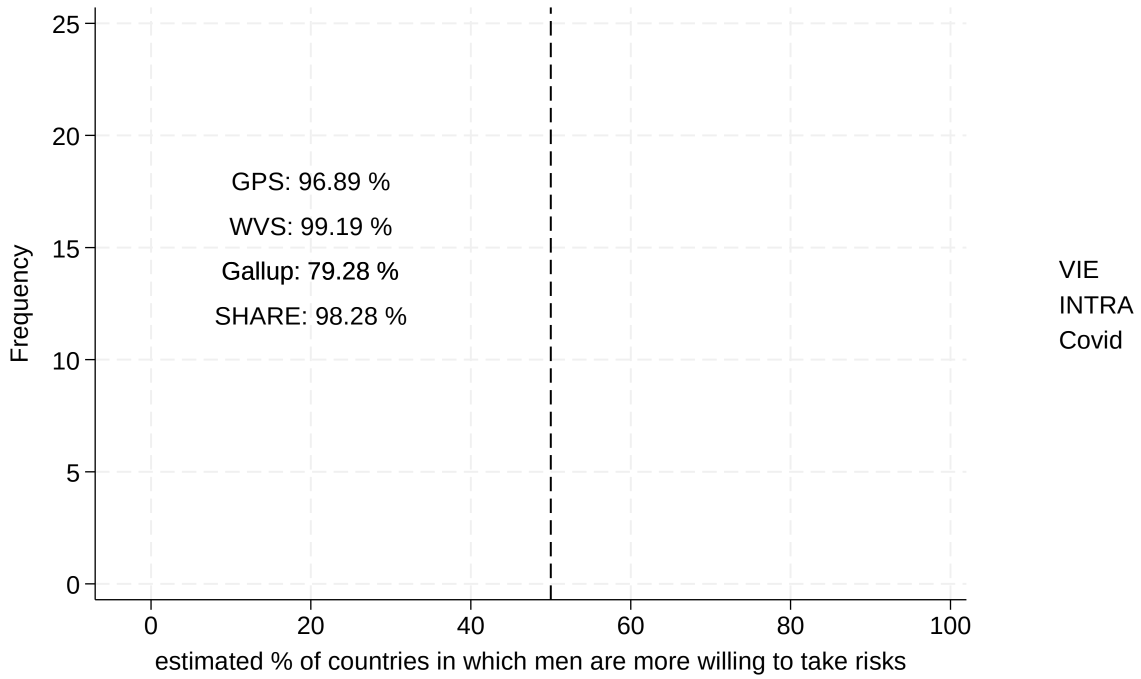


Figure 8 shows the estimated share of countries for which men are more willing to take risks than women for all 64 measures. For 83% of the measures, men are more willing to take risks in more than half of the countries. For 45% of the measure, men are on average more risk seeking in more than 90% of the countries.

At a first glance, Figures 7 and 8 suggest that while there is meaningful heterogeneity between the measures, men are generally more willing to take risks. However, this conclusion may be an artefact of the measures I have access to. For example, 54 are from lottery measures, 9 are from people's subjective self-assessments, and 1 is a hybrid measure (GPS). Among the 54 lotteries, 30 are in the gain domain and only 24 are in the loss domain, 38 are risk lotteries and only 16 are uncertainty lotteries. The appearance that men are generally more willing to take risks may be a result of, for example, disproportionately relying on risk lottery measures in the gain domain.

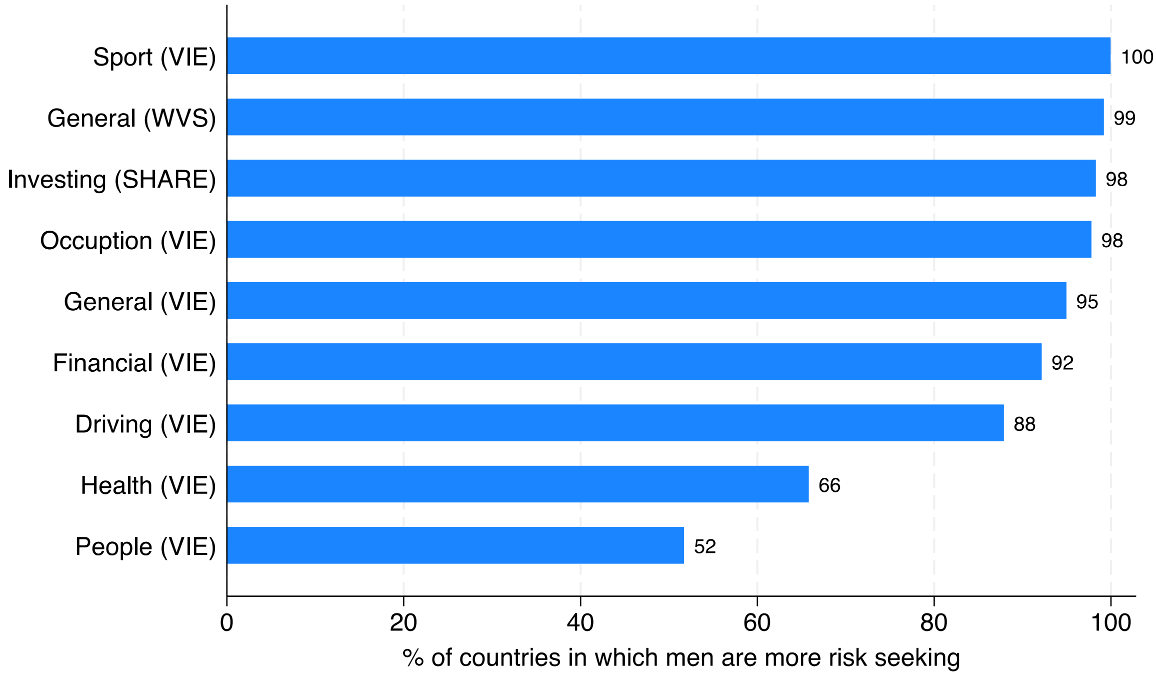
Figure 8: Estimated share of countries in which men are more willing to take risks, by measure



5.4. Heterogeneity by measure and country characteristics

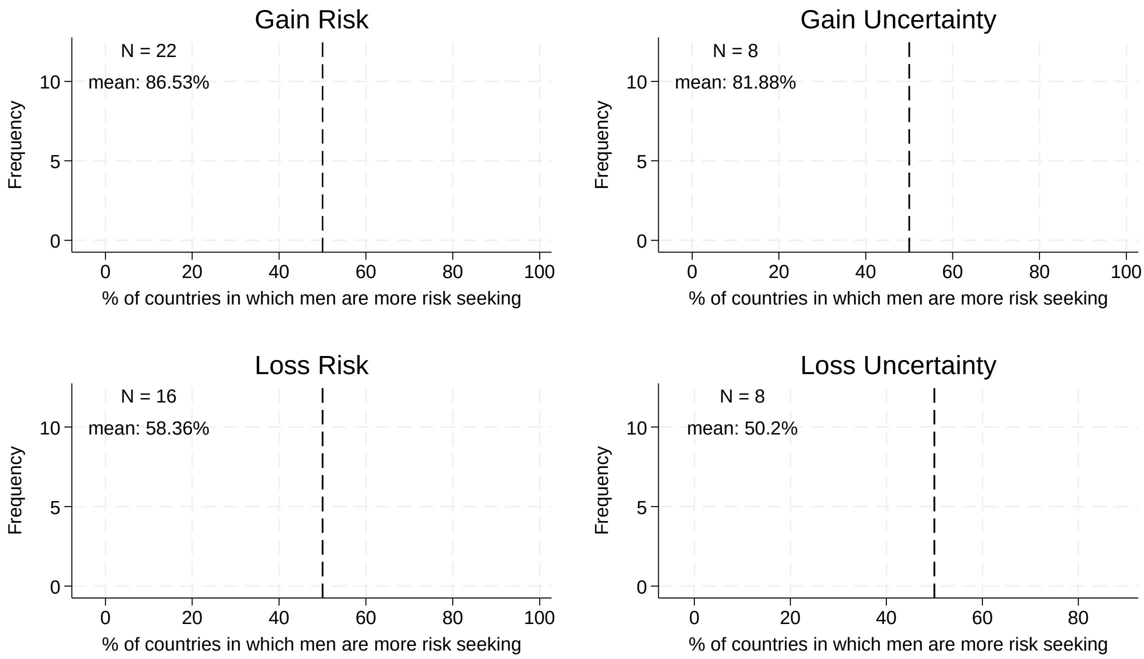
Figure 9 shows the estimated percentage of countries in which men are more willing to take risks for all 9 self-assessment measures. For all of those measures, my estimates suggest than men are more willing to take risks in more than half of the countries. For the two measures assessing people's general risk attitudes, men are more willing to take risks in more than 95% of the covered countries. We see similarly high figures for measures covering sport, investing, occupation, finance, and driving. Two notable exceptions are health and people: in those domains, men are more willing to take risks in just over half of the countries.

Figure 9: Estimated % of countries in which men are more willing to take risks (Survey measures)



Focussing on the 54 lottery measures, Figure 10 shows the percentage of countries in which men are more willing to take risks in four different categories: risk lotteries in the gain domain, uncertainty lotteries in the gain domain, risk lotteries in the gain domain, and uncertainty lotteries in the loss domain. We see no meaningful differences in results between risk and uncertainty lotteries. However, whether lotteries are in the gain or loss domain matters a lot. In the gain domain, most measures suggest that men are more willing to take risks in the majority of countries. In the loss domain, sex-differences are much less obvious. According to 15 of the 24 measures, men are more willing to take risks in more countries. According to 9 measures, women are more willing to take risks in more countries.

Figure 10: Heterogeneity by gain vs loss and risk vs uncertainty



I also test which measure characteristics or country characteristics matter with bivariate meta regressions of the coefficients of interest on binary indicators characteristics as independent variables. Table 3 reveals that lottery measures show 0.125 SD lower sex-differences in risk attitudes than survey measures (p-value <0.001), gain lotteries show 0.113 SD sex-differences than loss lotteries (p-value <0.001), and not meaningful sex-differences between risk and uncertainty lotteries. We see no statistically significant difference between WEIRD and non-WEIRD countries. However, sex-differences in risk attitudes appear larger in countries with an above median Human Development Index, above median GDP per capita. Sex-differences appear smaller in countries who score above the median on the UN's Gender Inequality Index. In sum, sex-differences in risk attitudes are for gain lotteries and in rich, developed, and gender-equal countries.

Table 3: Heterogeneity by measure and country characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lottery measure (vs. survey)	-0.125***						
	(0.011)						
gain lottery (vs. loss lottery)		0.113***					
		(0.010)					
uncertainty lottery (vs. risk lottery)			0.004				

			(0.012)				
WEIRD country				0.012			
				(0.010)			
above median HDI					0.029***		
					(0.010)		
above median GDP pc						0.036***	
						(0.010)	
above median GII							-0.032***
							(0.010)
Constant	0.203***	0.011	0.077***	0.105***	0.093***	0.089***	0.120***
	(0.010)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
Observations	1,904	1,594	1,594	1,980	1,974	1,974	1,961

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6. Conclusion

Sex-differences in risk attitudes are often treated as if they are universal. To see if this treatment is justified, I have analysed 1,980 country-level sex-differences in risk attitudes from 64 different measures covering more than 500,00 people in 120 countries. None of these estimates has been filtered through the publication process and I have used meta-analysis methods to correct for sampling error.

Taken together, my results suggest that sex-differences in risk attitudes are not universal but depend on the domain. On balance, we would expect men to be more willing to take risks. All kinds of measures either suggest that men are more willing to take risks or show no clear sex differences. No kind of measure suggests that women are more willing to take risks. However, the most frequently used risk measures in the literature–general self-assessments and lotteries in the loss domain–are also the kind of measures for which sex-differences in risk attitudes are most pronounced. This fact likely causes literature reviews, and the average reader to overestimate the prevalence and magnitude of sex-differences in risk attitudes. If it would be customary to measure risk attitudes with loss lotteries, we would likely have much less confidence in men being generally more willing to take risks.

References

Angrist, Joshua D., and Peter Hull. 2023. "Instrumental variables methods reconcile intention-to-screen effects across pragmatic cancer screening trials." *Proceedings of*

the National Academy of Sciences 120 (51):e2311556120. doi: doi:10.1073/pnas.2311556120.
Angrist, Noam, and Rachael Meager. 2023. "Implementation Matters: Generalizing Treatment Effects in Education." *Available at SSRN 4487496*.
Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, Sabrina Zuber, and on behalf of the SHARE Central Coordination Team. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4):992-1001. doi: 10.1093/ije/dyt088.
Bouchouicha, Ranoua, and Ferdinand M Vieider. 2019. "Growth, entrepreneurship, and risk-tolerance: a risk-income paradox." *Journal of Economic Growth* 24:257-282.
Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Hung HV Nguyen, Muna Adem, Jule Adriaans, Amalia Alvarez-Benjumea, Henrik K Andersen, Daniel Auer, and Flavio Azevedo. 2022. "Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty." *Proceedings of the National Academy of Sciences* 119 (44):e2203150119.
Brodeur, Abel, Scott E Carrell, David N Figlio, and Lester R Lusher. 2023. Unpacking p-hacking and publication bias. National Bureau of Economic Research.
Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star wars: The empirics strike back." *American Economic Journal: Applied Economics* 8 (1):1-32.
Byrnes, James P., David C. Miller, and William D. Schafer. 1999. "Gender differences in risk taking: A meta-analysis." *Psychological Bulletin* 125 (3):367-383. doi: https://doi.org/10.1037/0033-2909.125.3.367.
Charness, Gary, and Uri Gneezy. 2012. "Strong evidence for gender differences in risk taking." *Journal of economic behavior & organization* 83 (1):50-58.
Croson, Rachel, and Uri Gneezy. 2009. "Gender differences in preferences." *Journal of Economic literature* 47 (2):448-474.
de Gendre, Alexandra, Jan Feld, Nicolás Salamanca, and Ulf Zölitz. 2023. "Same-sex role model effects in education." *Working paper series/Department of Economics* (438).
Dekel, Eddie, and Suzanne Scotchmer. 1999. "On the evolution of attitudes towards risk in winner-take-all games." *Journal of Economic Theory* 87 (1):125-143.
Denzin, Norman K. 2015. "Triangulation." In *The Blackwell Encyclopedia of Sociology*.
Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the european economic association* 9 (3):522-550.
Eckel, Catherine C, and Philip J Grossman. 2002. "Sex differences and statistical stereotyping in attitudes toward financial risk." *Evolution and human behavior* 23 (4):281-295.
Eckel, Catherine C, and Philip J Grossman. 2008. "Men, women and risk aversion: Experimental evidence." *Handbook of experimental economics results* 1:1061-1073.
Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global evidence on economic preferences." *The Quarterly Journal of Economics* 133 (4):1645-1692.
Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *Management Science* 69 (4):1935-1950.
Filippin, Antonio, and Paolo Crosetto. 2016. "A reconsideration of gender differences in risk attitudes." *Management Science* 62 (11):3138-3160.

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk preference shares the psychometric structure of major psychological traits." *Science advances* 3 (10):e1701381.

García, Jorge Luis, and James J. Heckman. 2023. "Parenting Promotes Social Mobility Within and Across Generations." *Annual Review of Economics* 15 (1):349-388. doi: 10.1146/annurev-economics-021423-031905.

Gneezy, Uri, and Jan Potters. 1997. "An experiment on risk taking and evaluation periods." *The quarterly journal of economics* 112 (2):631-645.

Hagger, Martin S, Nikos LD Chatzisarantis, Hugo Alberts, Calvin O Anggono, Cedric Batailler, Angela R Birt, Ralf Brand, Mark J Brandt, Gene Brewer, and Sabrina Bruyneel. 2016. "A multilab preregistered replication of the ego-depletion effect." *Perspectives on Psychological Science* 11 (4):546-573.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "Cooperation, reciprocity and punishment in fifteen small-scale societies." *American Economic Review* 91 (2):73-78.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The weirdest people in the world?" *Behavioral and Brain Sciences* 33 (2-3):61-83. doi: 10.1017/S0140525X0999152X.

Holt, Charles A, and Susan K Laury. 2002. "Risk aversion and incentive effects." *American economic review* 92 (5):1644-1655.

Hoogeveen, Suzanne, Alexandra Sarafoglou, Balazs Aczel, Yonathan Aditya, Alexandra J Alayan, Peter J Allen, Sacha Altay, Shilaan Alzahawi, Yulmaida Amir, and Francis-Vincent Anthony. 2023. "A many-analysts approach to the relation between religiosity and well-being." *Religion, Brain & Behavior* 13 (3):237-283.

Huntington-Klein, Nick, Andreu Arenas, Emily Beam, Marco Bertoni, Jeffrey R Bloem, Pralhad Burli, Naibin Chen, Paul Grieco, Godwin Ekpe, and Todd Pugatch. 2021. "The influence of hidden researcher decisions in applied microeconomics." *Economic Inquiry* 59 (3):944-960.

Jackson, C Kirabo, and Claire L Mackevicius. 2023. "What impacts can we expect from school spending policy? Evidence from evaluations in the US." *American Economic Journal: Applied Economics*.

Kachelmeier, Steven J, and Mohamed Shehata. 1992. "Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China." *The American economic review*:1120-1141.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2):263-292.

Klein, Richard A, Michelangelo Vianello, Fred Hasselman, Byron G Adams, Reginald B Adams Jr, Sinan Alper, Mark Aveyard, Jordan R Axt, Mayowa T Babalola, and Štěpán Bahník. 2018. "Many Labs 2: Investigating variation in replicability across samples and settings." *Advances in Methods and Practices in Psychological Science* 1 (4):443-490.

Kvarven, Amanda, Eirik Strømland, and Magnus Johannesson. 2020. "Comparing meta-analyses and preregistered multiple-laboratory replication projects." *Nature Human Behaviour* 4 (4):423-434.

Max Roser, Cameron Appel, and Hannah Ritchie. 2021. "Human Height." *Our World in Data* 2024 (January 9).

Meager, Rachael. 2019. "Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments." *American Economic Journal: Applied Economics* 11 (1):57-91.

Nair, Gopindra S, and Chandra R Bhat. 2021. "Sharing the road with autonomous vehicles: Perceived safety and regulatory preferences." *Transportation research part C: emerging technologies* 122:102885.
Niederle, Muriel. 2017. "8. Gender." In *The Handbook of Experimental Economics, Volume 2*, edited by H. Kagel John and E. Roth Alvin, 481-562. Princeton: Princeton University Press.
Pedroni, Andreas, Renato Frey, Adrian Bruhin, Gilles Dutilh, Ralph Hertwig, and Jörg Rieskamp. 2017. "The risk elicitation puzzle." *Nature Human Behaviour* 1 (11):803-809.
Rieger, Marc Oliver, Mei Wang, and Thorsten Hens. 2015. "Risk preferences around the world." *Management Science* 61 (3):637-648.
Riley, Richard D, Paul C Lambert, and Ghada Abo-Zaid. 2010. "Meta-analysis of individual participant data: rationale, conduct, and reporting." *Bmj* 340.
Ruggeri, Kai, Amma Panin, Milica Vdovic, Bojana Većkalov, Nazeer Abdul-Salaam, Jascha Achterberg, Carla Akil, Jolly Amatya, Kanchan Amatya, and Thomas Lind Andersen. 2022. "The globalizability of temporal discounting." *Nature Human Behaviour* 6 (10):1386-1397.
Sapienza, Paola, Luigi Zingales, and Dario Maestripieri. 2009. "Gender differences in financial risk aversion and career choices are affected by testosterone." *Proceedings of the National Academy of Sciences* 106 (36):15268-15273.
Schwartz, Shalom H. 2012. "An overview of the Schwartz theory of basic values." *Online readings in Psychology and Culture* 2 (1):11.
Seidler, Anna Lene, Mason Aberoumand, Kylie E Hunter, Angie Barba, Sol Libesman, Jonathan G Williams, Nipun Shrestha, Jannik Aagerup, James X Sotiropoulos, and Alan A Montgomery. 2023. "Deferred cord clamping, cord milking, and immediate cord clamping at preterm birth: a systematic review and individual participant data meta-analysis." *The Lancet*.
Steinert, Janina I., Henrike Sternberg, Hannah Prince, Barbara Fasolo, Matteo M. Galizzi, Tim Büthe, and Giuseppe A. Veltri. 2022. "COVID-19 vaccine hesitancy in eight European countries: Prevalence, determinants, and heterogeneity." *Science Advances* 8 (17):eabm9825. doi: doi:10.1126/sciadv.abm9825.
Tversky, Amos, and Daniel Kahneman. 1981. "The framing of decisions and the psychology of choice." *science* 211 (4481):453-458.
Vardy, Tom, Cristina Moya, Caitlyn D Placek, Coren L Apicella, Alexander Bolyanatz, Emma Cohen, Carla Handley, Eva Kundtová Klocová, Carolyn Lesorogol, and Sarah Mathew. 2022. "The religiosity gender gap in 14 diverse societies." *Religion, Brain & Behavior* 12 (1-2):18-37.
Vieider, Ferdinand M., Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk, and Peter Martinsson. 2015. "Common Components of Risk and Uncertainty Attitudes Across Contexts and Domains: Evidence from 30 Countries." *Journal of the European Economic Association* 13 (3):421-452. doi: 10.1111/jeea.12102.
Weghorst, Keith R, and Staffan I Lindberg. 2013. "What drives the swing voter in Africa?" *American Journal of Political Science* 57 (3):717-734.
Yamada, Yuki, Dominik-Borna Ćepulić, Tao Coll-Martín, Stéphane Debove, Guillaume Gautreau, Hyemin Han, Jesper Rasmussen, Thao P Tran, and Giovanni A Travaglino. 2021. "COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak." *Scientific data* 8 (1):3.