Data Dilemmas:

The Science and Politics of Communicating Uncertainty in Human Rights Information

by

Tamy Guberek

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2019

Doctoral Committee:

Professor Kentaro Toyama, Chair
Professor Jay D. Aronson, Carnegie Mellon University
Professor Tanya Rosenblat
Professor Sonja Starr

Tamy Guberek

tamyg@umich.edu

ORCID iD: 0000-0001-5291-6713

Dedication

To Pojo, who deserves a Ph.D. in love and support.

Acknowledgments

First, I would like to thank all of the committed human rights advocates who generously shared their experiences and perceptions with me. I admire their professionalism and their courage as they face daily challenges in the fight for human rights and justice.

Heartfelt thanks to my incomparable advisor and dissertation committee chair, Kentaro Toyama. Kentaro has shown me the true meaning of mentorship. He has been caring and boundlessly supportive to me not only during my doctoral milestones, but through many life experiences over the past four years. I am deeply grateful that since the day we met, he has encouraged me to purse the questions I really care about, believing in me and supporting me every step of the way.

I would also like to thank the members of my dissertation committee, Tanya Rosenblat, Jay Aronson, and Sonja Starr for their continued guidance and constructive feedback as I imagined and carried out this research project. Tanya has been a role model to me – balancing research, teaching, and motherhood with grace. She has patiently helped me appreciate the iterative nature of experimental research. Thanks to Jay for always being available on the other side of a skype or phone call, to talk about rough ideas, research challenges, or life decisions. I feel privileged to have learned from his critical theoretical lens and sincere commitment to human rights. I greatly appreciate Sonja's contributions and support to this project, which is vastly better thanks to her sharp commentary each step of the way. I feel honored to have had such brilliant scholars accompanying me on my journey to becoming a better researcher.

Thank you to Alexa Patrick-Rodriguez, who provided valuable research support to me during the 2018-2019 year. She engaged in this project with perceptiveness and commitment, bringing insights that I would have otherwise missed.

It has been my great fortune to be part of a vibrant research community. Sincere thanks to the UMSI faculty with whom I have had the privilege to work and who have taught me to be a rigorous scholar and teacher: Florian Schaub, Margaret Hedstrom, David Wallace, John J. King, Cliff Lampe, and Erin Krupka. I am also grateful to have been part of the team at SPI Lab (Security and Privacy Interaction Lab) – all of its members have not only been inspiring research collaborators, but dear and supportive friends. Thank you Alison McDonald, Sylvia Simioni, Allan Martell, Abraham Mhadli and Yixin Zou. Also, huge thanks to other colleagues who have supported me in many ways through this Ph.D. program: Mohamed Abbadi, Tawfiq Ammari, Lindsay Blackwell, Melissa Chalmers, Priyank Chandra, Rebecca Frank, and Carol Moser. My warmest thanks to Joy Wilke, who has been a wonderful friend and dissertation writing partner. Her bright smile, generous advice, and constant encouragement got me through the final stages of the dissertation process.

Thank you also to the staff at the University of Michigan School of Information, who have helped me to navigate the doctoral program. I am particularly grateful to Veronica Falandino and Allison Sweet.

I am so fortunate to have incredibly supportive parents, Ingrid and Hertzl, who believe in me and accompany me through all of life's adventures. My sister, Shelly, who is always with me for a hard laugh, a desperate cry, or a simple honest moment. I am blessed to have Leah and Ilan, who brighten my each and every day with their picasos, giggles, and wonderings about the world. Thank you for being so understanding as I spent days and nights away from you to finish this dissertation. And finally, to Daniel, who's love, support, and patience know no bounds – thank you is not enough. You are my rock, my love, my life.

This research was funded in part by a Rackham Predoctoral Fellowship.

Table of Contents

Dedication .... **ii**

Acknowledgments .... **iii**

List of Tables .... **ix**

List of Figures .... **x**

List of Appendices .... **xi**

Abstract .... **xii**

Chapter 1 Introduction **1**

1.1 Research Questions and Methods 5

1.2 Findings and Contributions 6

Chapter 2 Literature and Theoretical Framework **9**

2.1 The Power of Numbers 9

2.2 The Multi-faceted Problems with Numbers 12

2.3 Data Uncertainty 14

2.4 Contending with Uncertain Data and Numbers 15

2.5 Transparency about Data 18

2.6 Communicating Uncertainty 21

2.7 Open Questions 24

Chapter 3 Research Methods **26**

3.1 Introduction to Mixed Methods Design ... 26
3.2 Qualitative Research ... 27
3.2.1 Population of Study ... 28
3.2.2 Analysis ... 33
3.3 Pilot Study of Sample of Human Rights Watch Reports ... 34
3.4 Survey Experiment ... 35
3.4.1 Experimental Design ... 35
3.4.2 Treatments ... 35
3.4.3 Vignettes ... 37
3.4.4 Treatments ... 41
3.4.5 Controls ... 43
3.4.6 Example Vignette and Treatment Language ... 46
3.4.7 Subject Pool and Sample Size ... 48
Chapter 4 Qualitative Interview Findings ... 50
4.1 Advocates emphasize value of numbers in advocacy ... 55
4.1.1 Advocates find that numbers compel attention ... 55
4.1.2 Advocates believe numbers offer complementary, descriptive value ... 59
4.1.3 Advocates also find numbers offer political "cover" ... 62
4.2 Advocates have a discerning sense of data limitations ... 63
4.3 Advocates contend with tension between numbers and data uncertainty ... 67
4.3.1 Advocates share consistent ideas about what constitutes "good enough" numbers ... 69
4.3.1.1 Four aspirations: empirical, temporary, conservative, and transparent ... 71
4.3.2 While advocates express transparency as an ideal, it is rarely realized ... 74
4.3.2.1 Methods sections are often high-level ... 75
4.3.2.2 Advocates describe deterrents and risks in presenting data limitations ... 78
4.3.2.3 Preliminary review of HRW reports ... 83

4.3.3 Methodological scrutiny is selective when advocates cite “authoritative numbers” . 84

4.4 Conclusion 88

Chapter 5 Experimental Survey Findings 90

5.1 Introduction 90

5.1.1 Sample Description 93

5.2 Main Findings 94

5.2.1 Decisions 94

5.2.1.1 The Power of Numbers 96

5.2.1.2 Effective and Ineffective Uncertainty Messages 97

5.2.2 Trust in Information Provider 101

5.3 Study Limitations 105

5.4 Conclusion 106

Chapter 6 Discussion and Conclusion 108

6.1 Key Findings 109

6.2 A Rigor-Pragmatism Continuum 116

6.3 Refining Appeals to Rigor, Transparency, and Credibility 123

6.4 Study Limitations and Future Research 126

6.5 Conclusion 129

References 142

List of Tables

Table 1: Interviewees 29

Table 2: Overview of Vignettes 39

Table 3: Primary Hypotheses 45

Table 4: C0 (No Number) vs. C1 (Number) Decision Means 96

Table 5: Decision as Dependent Variable 99

Table 6: Regressions for Perceptions of Trust 103

Table 7: Data uncertainty forces trade-offs on a rigor-pragmatism continuum for many issues involved in producing and presenting numbers 119

Table 8: Treatment language Vignette 1 136

Table 9: Treatment language Vignette 2 138

Table 10: Treatment language Vignette 3 140

List of Figures

Figure 1: Decision Mean Standardized relative to C1 95

Figure 2: Decision Means per Vignette 98

Figure 3: Perceptions of Trust in Information Provider, Standardized to C1 102

Figure 4: Perceptions of Trust in Information Provider 102

List of Appendices

Appendix A: Interview Protocol for Qualitative Study 131

Appendix B. Vignettes and Treatment Language 135

Abstract

Data and statistics about crime and human rights violations are incomplete and biased, yet numbers are in high demand. Advocates and policymakers often tally up available, yet partial data and present them as hard numbers to bring attention to abuses and to influence aid and accountability. As calls for transparency about data limitations increase, I ask two related questions: 1) How do human rights advocates think about the value of quantification and its associated uncertainty when using it to inform and influence audiences? 2) With respect to quantitative evidence about violence, crime or abuse, how do different presentations of data uncertainty affect decision outcomes? Using mixed methods – qualitative and experimental – this research teases out the political, behavioral and methodological challenges that advocates face as they collect, communicate, and deploy violence statistics in global and local human rights advocacy contexts.

Semi-structured interviews with twenty-eight frontline human rights advocates (focused on global, U.S., or Colombian issues) reveal that data uncertainty is an unavoidable reality in human rights work, and advocates are keenly aware of this. Advocates mostly share consistent and as-of-yet unrecognized ideas and practices about what could be called "good enough numbers" for advocacy. Central to these practices are pragmatic, yet principled tradeoffs that pull advocates away from strictly rigorous treatment of data and uncertainty. Transparency is a key issue that advocates somewhat reluctantly reduce in pragmatic considerations of benefits and risks.

The survey experiment employed three vignettes and four uncertainty messages, designed on the basis of science communication theory and human rights communication practices, to explore the impact of "being transparent" about data limitations. Responses from 970 college graduates confirm that 1) numbers have strong anchoring effects, but also show that 2) simple caveats about uncertainty do little to de-anchor decision-making. The research also finds novel evidence that 3) only one message type – called here "expert interception" – effectively drives people to account for uncertainty in their decisions (replicating earlier findings about communicating uncertainty in weather forecasting (Joslyn & LeClerc, 2013)). Finally, 4) while different studies suggest perceptions of trustworthiness of information providers may increase, or decrease, with different levels of uncertainty information, this study finds minimal fluctuation in source trust across any of the tested uncertainty messages.

Information providers face a clear choice in allowing numbers to speak for themselves or proactively mitigating bias through language – a choice that is inherently political. It appears that uncertainty is most effectively conveyed when communicators intercept the power of numbers to project "mechanical objectivity" with their expert knowledge about the data generation process and data limitations. A core theoretical contribution of this dissertation is the elucidation of a "rigor-pragmatism continuum" – a novel framework informed by the research findings. The continuum challenges the good-bad dichotomy that is common in critiques of human rights numbers and offers an alternative to support more nuanced analysis about how human rights advocates wrestle with using uncertain numbers. As a whole, this dissertation has wide-reaching implications for human rights and science communication scholarship and practice.

Chapter 1 Introduction

Collecting fact-based evidence to increase the likelihood of providing attention to atrocities, remedies for victims, acknowledgment of wrongdoing, as well as justice, reconciliation, and non-repetition, energizes much of the work in the human rights (HR) community today (Alston & Knuckey, 2016). In this context, scholars and advocates point out that quantitative data such as body counts, crime statistics, and data about abuses and abusers are in high demand to inform and influence decisions on behalf of vulnerable groups (Andreas & Greenhill, 2010; Cohen & Green, 2012; Langford & Fukuda-Parr, 2012; Merry, 2016; Root, 2016; Rosga & Satterthwaite, 2009; Satterthwaite & Simeone, 2016). Although the "information politics" of transnational organizations have long been central to the human rights advocacy movement (Keck & Sikkink, 1998), now more than ever, statistics, figures, and aggregate trends hold the promise – or perhaps just the veneer – of serving as powerful packages of objectivity in international, politicized, and adversarial contexts (Andreas & Greenhill, 2010; Merry, 2016; Porter, 1995). An inherent feature of the "seduction of quantification" is the elimination of uncertainty somewhere along the line from information production to consumption (Merry, 2016).

But the "seduction of quantification" faces limitations that are due, in part, to the inherent uncertainty in data about political violence, crime, and human rights abuses. These phenomena are hard to document for many reasons: perpetrators have strong incentives to eliminate evidence of their crimes, impede efforts to collect evidence, and hinder access to the evidence that does exist (Guberek & Hedstrom, 2017; Leiby, 2012). People and organizations working to document and collect such evidence may encounter silence from victims and witnesses (Roth, Guberek, & Green, 2011). In addition, data producers may themselves choose silence given the risks and high costs associated with collecting data such as testimonies, documents, and administrative records. In many cases, violence exceeds our capacity for documentation. In short, data and statistics about crime and mass violence rarely represent reality; they are almost always incomplete and biased in non-random ways that we cannot easily infer because ground truth (i.e., what actually happened) is almost never attainable (Davenport & Ball, 2002; Guzmán, Guberek, & Price, 2012; Kruger, Ball, Price, & Green, 2013; Landman & Godhes, 2013; Price & Ball, 2015b; Weidmann, 2016; Williams, Bowman, & Jung, 2016).

In human rights scholarship, the way statistics are produced and used has become a topic of critical inquiry. Due to their simplified form and the often unverified assumption that quantitative experts have produced the numbers, audiences tend to "mechanically" read numbers as objective (Porter, 1995). Porter aims to conceptually uncouple the conflation of objectivity from "truth," arguing instead that objectivity in numbers should be understood more as a procedural "distance from subjectivity" based on universal rules for the production of quantification. Merry finds that nonetheless the "seduction of quantification" is palpable in human rights governance forums as "the idea that indicators offer a particularly reliable form of truth" (p.26). The problem is exacerbated given that with the "turn to metrics," embedded uncertainty in data falls away somewhere along the line from information production to consumption (Langford & Fukuda-Parr, 2012; Merry, 2016; Rosga & Satterthwaite, 2009). Scholars from different disciplines, who employ a variety of research methods, are concerned with data quality and the relationship between data and the phenomena that they represent. This “data uncertainty” is often rendered invisible and effectively eliminated when observational information about complex social phenomena is reduced to numeric form. Given the inherent limitations to observing and documenting violence, crime and abuse, I focus here on this most initial form of uncertainty that enters the quantitative knowledge production pipeline.

Scholars are sending cautionary messages about bias in data, algorithms and statistics in the human rights field, and much beyond (Andreas & Greenhill, 2010; boyd & Crawford, 2012; Goldstein, 1986; O’Neil, 2016; T. B. Seybolt, Aronson, & Fischhoff, 2013; Tufekci, 2014). In the human rights realm, there is a growing concern that overly confident quantitative claims and failure to present underlying data uncertainty to information consumers can lead to real consequences: undermine the credibility of organizations (Cohen & Green, 2012), amplify racial and socio-economic bias in criminal justice decision-making (Lum & Isaac, 2016; Starr, 2014), enable funneling of attention and resources to only the most visible problems, distort scholarly knowledge production (Landman & Godhes, 2013), and privilege certain narratives over others (Kruger et al., 2013; Price & Ball, 2015b; Reydams, 2016; Weidmann, 2016). Overall, these issues can run counter to the historical clarification and broader truth-telling project of the human rights community.

Due to these concerns, scholars from across disciplines are calling for more transparent communication of data limitations to audiences, especially when important decisions are at stake (boyd & Crawford, 2012; Greenhill, 2010; O’Neil, 2016; Root, 2016; Satterthwaite & Simeone, 2016; Tufekci, 2014). Given the expertise, cost, and difficulties often involved in calculating statistical bounds of uncertainty (Fariss, 2014; Kruger & Lum, 2015; Price & Ball, 2015b; T. B. Seybolt et al., 2013), an alternative recommendation to practitioners has been to “confront limitations and bias [in data] through language” (Root, 2016, p. 364). The underlying idea here is that by expressing uncertainty, the communicator offers complementary information to audiences that can buffer against false precision in interpreting quantitative information. This call for transparency naturally aligns with the truth-telling agenda of the human rights community.

However, at present, we know little about what these tensions mean for human rights practitioners, nor how they view or manage them in their current communication practices. Scholarly commentary often portrays HR advocates with these data dilemmas as using "bad data" liberally to advance their causes (Best, 2012; Cohen & Green, 2012). Disclosure about data quality appears to be minimal: Simone and Satterthwaite (2016) reviewed the methodology sections in ten years of advocacy reports by two leading human rights NGOs and find that descriptions on the difficulty of data collection were "unusually abbreviated," and discussed in less than one third of all the reports produced in 2010 (p. 325). Heinzelman & Meier (2013) suggest that any acknowledgement of data limitations around numbers tends to be in the form of empty caveats, and Ball (2016) suggests that the notion that caveats help readers calibrate their consumption of numbers is a "fallacy."

Cognitive science and science communication scholarship reveals that communicating uncertainty effectively in any domain is not straightforward. Research from disparate scientific fields provide scattered insights on the challenges of communicating uncertainty. Cognitive studies show that better decision outcomes emerge from numerical expressions of uncertainty than linguistic ones (Budescu, Por, & Broomell, 2012; Budescu & Wallsten, 1987; Joslyn & LeClerc, 2013; K. A. Martire, Kemp, Watkins, Sayle, & Newell, 2013); that making explicit recommendations to decision-makers in the face of uncertainty is likely to lead to improved outcomes (Joslyn & LeClerc, 2013; Joslyn, Nadav-Greenberg, & Nichols, 2009); that "hedges" can increase perceptions of credibility of information providers (Jensen, 2008), but can also weaken the perceived value of evidence in court (Kadane & Koehler, 2018; K. Martire, Kemp, Sayle, & Newell, 2014).

Yet even in these other domains, the findings on communicating uncertainty are isolated and sometimes contradictory, and the relative degrees to which uncertainty messages affect decision-making and perceptions of credibility in the sources remain unknown. Therefore, despite the growing attention to data problems and speculation about its consequences, there is little hard evidence that communicating data uncertainty sways decisions one way or the other. Until now, in human rights scholarship, while there is a growing desire for more explicit communication of data uncertainty, how to understand the science and politics of such communication remains an open challenge.

1.1 Research Questions and Methods

While we know some about the appeal of quantification and the problem of uncertainty, this dissertation research aimed to answer two open, overarching questions:

1. How do human rights practitioners think about the value and impact of quantitative evidence and its associated uncertainty when communicating it to inform and influence audiences?
2. When presented with quantitative evidence about violence, crime, or abuse, how do different presentations of data uncertainty affect decision outcomes?

To answer these questions, I conducted a mixed methods study. For the first question, I conducted a qualitative interview study. I interviewed frontline human rights practitioners who provide or receive quantitative 'facts and figures' about crime or violent events in their daily work. I explored their views on the value of data and quantitative information in achieving their goals; their perceptions of the limitations in the data they use to make numeric claims; their practices for navigating data uncertainty; the language they use (or not) to present data and uncertainty. In a complementary way, the qualitative research helped me understand the broader context and the feasibility of adopting different communication practices about data and uncertainty.

Next, I isolated and tested the impact of four theoretically distinct ways to communicate data uncertainty in typical human rights decision-making scenarios using an online experimental survey. Participants considered three vignettes and answered survey questions on decisions and trust perceptions, as well as various other outcomes measures. Each treatment consisted of a different type and way to communicate the provenance of the data and its limitations. The qualitative and experimental research were sequenced in such a way that allowed me to draw rich and complementary insights for the dissertation as a whole (Creswell & Clark, 2011). The survey experiment on communicating data uncertainty conducted here is the first of its kind in the human rights literature. Finally, it is worth noting that the overall study is also informed by over a decade of my own professional experience working in the human rights community, conducting data-based research and advocacy.

1.2 Findings and Contributions

Aiming to contribute to the growing scholarship taking a critical look at the use of human rights statistics, this research makes at least six empirical findings. First, I learn through my interviews with human rights advocates that data uncertainty is an unavoidable reality in human rights work, and human rights advocates are keenly aware of this. Second, human rights advocates are generally conscientious about wanting to avoid presenting or citing "suspect facts" (Cohen & Green, 2012, p. 446). One key way they do this is by making use of what I call "good enough numbers," which I find to be a shared set of pragmatic, yet principled choices that pull advocates away from strict rigorous treatment of data and uncertainty. Third, human rights advocates consider various political and cognitive risks of "being transparent" about data and methods. Transparency about methods and data uncertainty is a key issue that advocates somewhat reluctantly suggest gets reduced to minimal form in pragmatic considerations of benefits and risks. Most reason that minimal transparency aligns with audiences' expectations, thus poses little risk to their institutional credibility – an intuition bolstered by the experimental results which yield minimal fluctuation in the source trust measure across treatments and controls. Forth, also via the survey experiment, I find that the message type I call "expert interception" is the only one to consistently attenuate the strong anchoring effect of biased numbers. The message, informed by similar findings in scholarship on communicating uncertainty about weather forecasts, anticipates audiences' tendency to read numbers as falsely precise and warns them about how such an interpretation will impact the decision outcome. In essence, expert interceptive messages intentionally ask audiences to engage with the information providers' expertise about the data production instead of mechanically placing their trust in the numbers. Fifth, among most advocates included in the interview study, trust in institutional authority appears to matter more than quantitative expertise. Finally, the experimental findings make clear that information providers face a clear choice to be made between allowing a number to speak for itself or being proactive about mitigating bias. Ultimately, human rights advocates, or anyone using numbers they know to be uncertain, must contend with this inherently political choice.

Drawing on insights about the role of quantification in society from previous scholars (e.g., Jasanoff, 2014; March & Simon, 1958; Merry, 2016; Porter, 1995), I formulate a new model to analyze the practical trade-offs that advocates face when considering to use uncertain data to advance their work. While prevailing debates about numbers in human rights tend to judge uncertain numbers as “good” or “bad” (Andreas & Greenhill, 2010; Best, 2012), I believe a rigor-pragmatism continuum is better suited to analyzing how advocates regularly and conscientiously wrestle with when and how to use less-than-perfect numbers. Nonetheless, I find the key weakness to advocates’ rationale to be the context-specific temporality they see in “good enough” numbers for advocacy. The short-term utility of “good enough” numbers contrasts with the unavoidable fluidity, anchoring effect, long-term staying power, and reuse of numbers.

I hope this research can be of service to human rights practitioners. As they work to be influential in an increasingly data driven world, I appeal to posterity as one important reason why investing in rigorous practices and increased transparency about methods and data uncertainty is worthwhile. While valuing the complex tradeoffs that advocates make to advance their agendas, rigor can nonetheless serve as a North Star for the community.

Chapter 2 Literature and Theoretical Framework

This dissertation builds on insights from diverse fields including science and technology studies, human rights, and science communication. In this chapter, I first examine the theoretical literature explaining the source of power of numbers in society, as well as some critical work on how that power easily facilitates the misuse of numbers. I find some scholars make a stark distinction between "good" and "bad" statistics which I argue may be an unhelpful framing for human rights statistics that are plagued with inherent and unresolvable data uncertainty. I define data uncertainty as the set of issues that make data have a hard-to-specify relationship with reality. I then review the literature on the production and use of human rights statistics and the existing recommendations on how to handle inherent data uncertainty about hard-to-document events of crime, violence, and abuse. Taking seriously recommendations that call for more transparent communication of data uncertainty, I analyze theories and empirical research from cognitive science and science communication fields about the ways and impacts of communicating uncertainty. I close with a synthesis of open questions related to communicating uncertainty about human rights violations data – open questions that motivate and inform this dissertation research.

2.1 The Power of Numbers

Quantification is an attractive form of communication in public life (Porter, 1995). In his seminal book *Trust in Numbers*, Theodore Porter theorizes that the power of numbers is based less on their accuracy and more on their form: they conveniently summarize complex information in adherence to universal, standardized rules that are recognizable across time and space. For this reason, he argues, numbers serve as communication ‘technologies of distance,’ whereby audiences need not rely on the subjective authority of the information provider and can instead rely on the authority of the universal rules that underlie the process of producing numbers. Number are seen to be “objective” not because of their relationship with truth in nature, but because of their distance from the subjective opinion of the information provider. Ultimately, numbers project “mechanical objectivity,” as it is assumed that experts rigorously follow the rules that ensure a check on subjective opinion and personal bias (even though, he says, “the rigor and uniformity nearly disappear in relatively private and informal settings” (ix). Rigor here is the idea that rules for proper quantification are judiciously followed and enforced by disciplinary peers (Jasanoff, 2014; Marquart, 2017). Overall, numbers serve as an information consensus where more proximate social and political connections are lacking – allowing them to transcend gulfs of trust among people and communities. By virtue of their simplified, unambiguous form and utility at a distance, numbers also afford incontestability, i.e., once produced, the compressed package of mechanical objectivity is difficult to reverse engineer and becomes relatively stable in its authority. As such, numbers have been found especially valuable to support decision-making in otherwise politicized environments (Porter, 1995, p. ix). Empirical scholars have confirmed quantification to command significant authority in fields such as criminal justice (Hannah-Moffat, 2013; Starr, 2014), law enforcement (Brayne, 2017), global governance (Merry, Davis, & Kingsbury, 2015), and human rights (Merry, 2016; Root, 2016; Rosga & Satterthwaite, 2009).

Keck and Sikkink (1998) define 'information politics' in advocacy work as "the ability to quickly and credibly generate politically usable information and move it to where it will have the most impact" (p.16). Scholars emphasize that for this reason, numbers are in high demand in the human rights community. With the "turn to metrics," advocates and policymakers feel an increased pressure to produce numbers. Some scholars have been studying how practitioners work with this pressure (Cohen & Green, 2012; Langford & Fukuda-Parr, 2012; Merry, 2016; Rosga & Satterthwaite, 2009). For example, Cohen and Green (2012) believe numbers have taken on added prominence in recent years as human rights advocates rely less on targeting select bad actors for traditional "naming and shaming" campaigns and more on emphasizing the gravity of human rights crises to multiple audiences. Merry (2016) depicts how advocates and policymakers working to measure on sexual violence and human trafficking team up with technical experts to produce and provide numbers in global governance forums like the United Nations and the U.S. State Department. She highlights a kind of pragmatism at work: despite contested political debates to determine the categories to classify these acts and the recognition of inherent uncertainty, the imperfection and messiness in the available information get washed away when indicators are produced and used. Root (2016) explains that human rights advocates, who tend not to be quantitative experts and instead have relatively low "quantitative literacy," often feel pressure to provide numbers. From his experience as Quantitative Analyst at Human Rights Watch, he considers that advocates "will produce and publish estimates of 'dark figures' when there is sufficient data and reasoning in calculation" (p.365). Emblematic of the pragmatic approximation to the data dilemmas, he writes "there is nothing inherently wrong with making an educated estimate when it is impossible to know the true extent of a phenomenon. It is essential that any estimates be treated with all the caveats required and not treated as fact" (p.367).

2.2 The Multi-faceted Problems with Numbers

Ultimately, trust in numbers relies on trust in experts (and trust in experts relies on vetting by disciplinary communities). Scholars concerned with veneers of objectivity inherently worry that the public relies too much on assumptions and appearances of judicious rule-following by experts even when those are not verifiable. As the demand and supply of numbers used to inform and influence public spheres has grown, scholars worry that "mechanical objectivity" leads to numbers being read with unquestioning trust, thus facilitating the propagation of "bad" numbers and misuse with little deterrence (Andreas & Greenhill, 2010; Best, 2012). In his influential book "Damned Lies and Statistics," Joel Best outlines how to distinguish "good" and "bad" statistics about social problems, a challenge complicated by what he calls the "public" and "hidden" power of statistics. The "pubic power" of statistics is that given the simplicity of their form, they "convert complicated social problems into more easily understood estimates, percentages, and rates." Their "hidden power" is that they are used to support views. Together, their simple, unambiguous form and uncritical consumption, he argues, makes it easy for statistics to be used by activists as "ammunition for political struggles." In their edited volume, Andreas and Greenhill (2010) include several case studies of the misuse of unfounded statistics in the international relations realm, citing the incentive for attention as the main driver of this practice.

Best describes the multiple ways activists produce “bad statistics.” First, he writes, “bad statistics” can arise from guessing without data or from estimates with flawed data. Simple guesses tend to be intuitive, qualitative, or political assessments that a phenomenon is severe and widespread. “Bad numbers” can also be data-based and still be fraught with social, political, and epistemological challenges. He outlines three main data problems: 1) The definitions for what is to be counted are often vague – too broad or too narrow, leading to counts that include false positives or false negatives. 2) Counting rules and measurement instruments are manipulatable and opaque. 3) Data samples can be unrepresentative, but are nonetheless used to make inferences and generalizations. The issues he raises as constituting “bad” statistics include a wide range – from guesses to various data problems. The data problems are especially difficult to deal with for data about complex social problems like human rights violations.

Best goes on to describe other classes of “bad statistics” which stem from innumeracy or deliberate manipulation. These can take forms such as inappropriate generalization, misrepresenting the meaning of otherwise “good” statistics, garbling complex statistics, or making inappropriate comparisons. Typical consumption of numbers based on trust (trust in numbers which bakes in trust in experts, i.e., Porter) rather than on verification enables the unchecked circulation of such statistics. Echoing some of the issues raised by Best, Cohen and Hoover Green (2012) describe the misuse of statistics by advocates who are quick to publish unverified and exaggerated claims, such as the claim that 75% of women experienced sexual violence during Liberia’s civil war between 1989-2003. They hypothesize more broadly that human rights advocacy organizations face “dueling incentives” when making quantitative claims: the need to impel “drama” in the short-term (for which there is the strong temptation to guess and exaggerate, i.e., misuse statistics in Best’s vocabulary) versus the need to maintain credibility in the long-term. Interestingly, Slovic et al. (2013) find psychological numbing effects when people face increasingly high numbers of mass atrocities, ultimately weakening their strength at compelling action.

All the issues raised in Best's "good vs. bad" framework are important. However, they include a heterogenous mix of issues – from unsolvable social, technical and epistemological challenges with data to wholly unfounded numbers being naively or intentionally propagated. In my view, putting all these issues in the same problematic bucket may not be the most helpful way to understand and address some of the underlying tensions with the inherently partial and biased nature of data about human rights violations. In this study, I uncouple the largely unsolvable data problems from the politics that may incentivize misuse. I start with the basis that numbers about human rights violations are usually, if not always uncertain. I will focus on analyzing how advocates manage the tension between uncertainty inherent in data and the politics of human rights advocacy.

2.3 Data Uncertainty

On the challenges of producing reliable data about human rights violations (HRVs), there is an ever-increasing body of empirical work demonstrating how HRVs are extremely difficult to document and measure: perpetrators try to conceal abuse, violations are hard to observe, victims comprise a "hidden" population, and different logics of access, trust, changing and limited resources. (Ball & Price, 2018; Cohen & Green, 2012; Davenport & Ball, 2002; Fariss, 2014; Guzmán et al., 2012; Hall & Stahl, 2008; Kruger et al., 2013; Landman & Godhes, 2013; Lum & Isaac, 2016; Roth et al., 2011; Weidmann, 2016). For any or all of these reasons, data tend to be biased towards what is known of groups who already tend to be disproportionately visible. Unrecognized bias threatens inferential endeavors (Eckhouse, Lum, Conti-Cook, & Ciccolini, 2019; Price & Ball, 2015a; T. B. Seybolt et al., 2013; Weidmann, 2016). In addition, there are the politics of defining, classifying and counting such complex phenomena (Coomans, Grunfeld, & Kamminga, 2009; Merry, 2016; Rosga & Satterthwaite, 2009).

The inevitable challenges in the process of attempting to measure HRVs means that almost all available datasets are flawed. Simple statistics on the basis of such data constitute what Best deems "bad statistics." However, "bad statistics" may be too harsh a judgment, especially because there is little that can be done to change or improve such data. Even some of the most tenacious documenting efforts are plagued with bias (Price & Ball, 2015a).

Rather than constituting "bad" data, one might consider HRVs data plagued by definitional, measurement, and sample limitations as having *an uncertain relationship with the reality they are used to portray*. Some scholars have called the uncertainty due to imperfect data "second-order uncertainty," or simply ambiguity, referring to "uncertainty created by 'the amount, type, reliability, and 'unanimity' of information and giving rise to one's degree of 'confidence' in the estimate…" (Dieckmann, Mauro, & Slovic, 2010; Ellsberg, 1961). In this dissertation, I use the term 'data uncertainty' to refer specifically to the set of issues that result in imperfect data that, despite their limitations, are nonetheless used to make formal or informal inferences about the world. I use the term "uncertain numbers," as distinct from the overly-broad term "bad numbers" to refer to numbers plagued by data uncertainty.

2.4 Contending with Uncertain Data and Numbers

Some of the scholars mentioned above make appeals to information producers and consumers about what to do in the face of overarching "bad statistics." For example, Best appeals to information consumers to "become better judges of the numbers we encounter" (p.13). He urges people to evaluate who, why, and how the numbers were produced. Evaluating the "who" asks consumers to verify that producers of numbers have expertise worth trusting. Evaluating the "why" asks consumers to look for "clues to the motivation" of the number provider (p.24). Evaluating the "how" asks consumers to look at the data and methods that produce the statistics, often requiring some minimal numeracy criteria. Overall, he asks the public to avoid being "awestruck," "naïve," or "cynical" consumers of social statistics, but to be "critical," i.e., to expend some thought to distinguish between good and bad statistics. Andreas and Greenhill (2010) make similar appeals, arguing it "behooves consumers to think harder about sources of data, the conclusions they draw from these data, and the assumptions on which they are predicated" (p. 4). However, given that the power of numbers relies on information consumers' "mechanical" assumptions of objectivity, such appeals may face an uphill battle. They essentially ask consumers to expend effort that will override the trust-in-numbers and trust-in-experts heuristics most people rely on. Such appeals also beckon people to seek information about the back-story to the production of numbers which is often in low supply (as we will see below).

In the human rights scholarship, experts have also been making appeals to information producers, largely based on their commitment to the rule-following required to make reliable quantitative claims. Critical scholars argue that using uncertain data to making quantitative claims could lead to consequences that jeopardize the broader truth project of the human rights community: they can amplify racial or demographic bias in predictive policing (Lum & Isaac, 2016), distort historical narratives (Ball, Asher, Sulmont, & Manrique, 2003), favor impunity for certain actors (Reydams, 2016), and possibly hurt the long-term strength and credibility of human rights organizations (Cohen & Green, 2012; Greenhill, 2010; Root, 2016; Satterthwaite & Simeone, 2016).

Critical human rights scholars have made what I see as three recommendations to practitioners using uncertain data and statistics. The first is that given data uncertainty, any effort to produce numbers should draw on statistical methodologies that account for the uncertainty of data and produce inferential estimates and calculated bounds of uncertainty – essentially to 'fix the data' with rigorous social science (Ball & Price, 2019; Fariss, 2014; T. B. Seybolt et al., 2013). 'Rigor' in this context usually means that quantitative claims are held to social scientific standards of quantitative research in the planning, data collection, analysis and reporting of findings, and it is in this methodological compliance and replicability that they derive their trustworthiness (Marquart, 2017; Satterthwaite & Simeone, 2016). However, advanced methods are only applicable under very specific conditions and require advance statistical expertise to calculate – expertise that is usually beyond the time, resources, or skills of typical a human rights project. The second recommendation is for practitioners to avoid using flawed data altogether to make descriptive, inferential, or predictive quantitative claims, i.e., to 'drop the data' (Andreas & Greenhill, 2010; Price & Ball, 2015b). As Andreas and Greenhill put it, advocates could just acknowledge that they "don't know" the magnitude and patterns of phenomena (p.278). Lastly – and perhaps the most broadly applicable yet least understood option – is for human rights practitioners to transparently inform audiences about the origins of the data and their inherent uncertainty (Ball & Price, 2018; Greenhill, 2010; Merry, 2016; Root, 2016; Satterthwaite & Simeone, 2016). For example, Satterthwaite and Simone (2016) write "[i]ncreased transparency about methods, research procedures, and limitations, as well as increased attention to the evidence base for conclusions and recommendations, will ultimately support the human rights fact-finding community by ensuring its credibility and reliability" (p.345). Sally Merry calls more specifically for "clear warnings about the limitation of indicators, the problems of missing or inadequate data, weak proxies, the generalization inherent in commensuration, and the loss of structural and systematic knowledge could improve indicator literacy" (p.217). Implicit in this last call is the idea that if information producers gave information consumers more information about the social, messy, and limited process of data production, consumers could better assess the quality, uncertainty, and reliability of the claims made on the basis of such data. In this way, transparency about data production and data uncertainty is seen as the path to ease some of the tension caught up in debates about the science and politics of using statistics to advance human rights.

Overall, in this research, I find that fixing and dropping numbers are not practical alternatives, which I will show in Chapter 4 Qualitative Interview Findings. On the basis of interviews with practitioners, I analyze how practitioners currently manage using uncertain data for advocacy goals. The main research question explored in this study is about the viability and impact of transparent communication of data uncertainty. To do so, in the next section, I explore the current literature on data transparency.

2.5 Transparency about Data

> "From philosophers concerned with the epistemological production of truth, through activists striving for government accountability, transparency has offered a way to see inside the truth of a system."
>
> Mike Ananny and Kate Crawford, 2018, p.974

Transparency entails that a system is made open to observation, usually for the purpose of giving more control to the viewer, ideally to enable accountability (Christensen & Cheney, 2015). In the ideal, "information is easily discernible and legible; that audiences are competent, involved, and able to comprehend" the information made visible (Christensen & Cheney, 2015, p. 74). In scientific communities, transparency means that the process behind any scientific claim is available for other scholars to carefully evaluate. It means that the credibility of the claims should not be based on assumed expertise and discretion, but on a reviewable and rigorous research process (Lupia & Elman, 2014).

In the human rights data context, the call for transparency is usually made as a benign petition aligned with best research practices. Nonetheless, the idea of accountability is inherent to such appeals – that transparency will allow audiences to better comprehend and ultimately judge the reliability of the information. With respect to data uncertainty, transparency is seen as pragmatic, compromising option to convey data limitations to audiences. A few scholars have taken a look at what transparency and disclosure of data limitations in human rights reports actually looks like. Satterthwaite and Simeone (2016) show how descriptions of data and methodological limitations are rare (although improving slightly over time) in reports by global human rights organizations like Amnesty International and Human Rights Watch. Overall, they find "unusually abbreviated time periods for investigations or especially difficult-to-access target populations" (p.325). In the related criminal justice literature, Hannah-Moffat (2013) worries that "…sources of data used to complete risk assessments are rarely disclosed or known to the court; only the risk score is provided" (p.285). According to Heinzelman and Meier (2013), if practitioners do address uncertainty, they usually do not go so far as to provide information that would lead to greater comprehension. Instead, they see hints of disclosure in the form of "caveats," carefully crafted short statements about the limitations of the data. Heinzelman and Meier posit that practitioners do this more in the spirit of protecting their credibility than of communicating uncertainty. Separately, Ball (2016) argues that the idea that caveats effectively communicate data limitations to audiences is a "fallacy."

Some scholars point to a fallacy of the ideal of transparency more broadly. Ananny and Crawford (2018) offer ten limits to the transparency idea; two are especially relevant here: the idea that transparency can “privilege seeing over understanding” and the illusion that transparency builds trust (p.980). In the human rights community, the calls for transparency are broad: some call for more information about research methods for fact-finding; others make specific appeals about communicating uncertainty in data. However, the information required and the implications of these two are distinct. We know from science communication literature that communicating uncertainty it is rarely done (as I will review below). In the specific context of human rights measurement, Merry (2016) argues that in the global human rights community, the “seduction of quantification, the idea that numerical data offer a particularly reliable form of truth,” is inherently linked to the elimination of uncertainty (p.26). She usefully evokes the idea of “uncertainty absorption” from March and Simon (1958) which “takes place when inferences are drawn from a body of evidence, and the inferences rather than the evidence itself, are then communicated” (p.165). As data get input, processed and output through the rule-making and number-producing pipeline, flawed or missing data, conceptual proxies, fuzzy definitions, the lack of commensurability and the political shaping of the categories for counting get pragmatically “stripped away” (Merry, 2016, p. 5). In theory then, communicating uncertainty that stems from the process of producing information may sit like an oxymoron next to quantitative information. In human rights and much beyond, communicating uncertainty in data or scientific findings is thought to come with risks. As Brian Root, Quantitative analyst at HRW puts it, “[i]n general, I believe practitioners are uncomfortable with uncertainty….because it threatens their strength in a human rights project rooted in accuracy, veracity and integrity – the use of numbers with some level of uncertainty is seen to endanger accuracy” (p.365). However, if the perceived persuasive power of numerical claims depend on methodological brevity or wholesale suppression of uncertainty (Lupia, 2013; Merry, 2016), and actors view uncertainty as a potential vulnerability (Andreas & Greenhill, 2010; Root, 2016), calls for communicating uncertainty may remain mute.

At present, we only know little about the extent to which human rights practitioners contend and possibly mitigate the tension between an inevitably uncertain supply of information about HRVs, and the demand for authoritative numbers. We know less about how calls to be more transparent about data resonate in complex and politized real-world contexts. And, in the human rights context, if we are to take seriously the calls for more transparency about data uncertainty, what would such expressions look like and what is the impact of conveying them to audiences? To begin considering how to answer these questions, in the next section I review theoretical and empirical insights from scholarship on communicating science and uncertainty to inform my research design.

2.6 Communicating Uncertainty

Until recently, there was a general consensus across domains that authors should shield their audiences from uncertainty. This was partly due to: 1) the well-established idea that cognitively, people are poor at processing and making decisions with statistics and uncertainty (Kahneman, 2011), 2) the idea that people make decisions using motivated reasoning and heuristics based on source trust (Kahan, 2011), and 3) the belief that uncertainty brings vulnerability in adversarial or politicized environments, where it can be taken out of context and used against its providers and their scientific findings (Fischhoff, 2012; Oreskes & Conway, 2010). More generally, the science communication literature finds little room for communicating uncertainty when attempting to be persuasive in politicized environments (Lupia, 2013). In fact, scholars find that in public policy domains, scientists tend to present results with "incredible certitude," shielding audiences from uncertainty in ways that they argue are detrimental for informed decision-making (Fischhoff, 2012; Manski, 2018).

Disparate insights from work on the use of forensic evidence in criminal cases, weather forecasting, and climate change research (among many other disciplines) suggest that different ways of expressing uncertainty can lead to varied outcomes. Crucially, the presenter of the information and the presentation format can have important effects. With respect to presentation formats, there is consistent evidence that numerical expressions of uncertainty (probabilities that specify both the location and degree of uncertainty with respect to some estimated reference point) are more reliably interpreted than linguistic expressions (Joslyn & LeClerc, 2013; K. Martire et al., 2014; Wallsten & Budescu, 1995). At the same time, there is evidence that people tend to be overconfident (e.g., ignoring confidence intervals and focusing on point estimates) even when calculated confidence intervals are provided (Correll & Gleicher, 2014; Tversky & Kahneman, 1974). Verbal expressions of uncertainty or "estimative language" (such as "unlikely", "good chance", and "probable") are interpreted with more inter-subjective and cross-context variation that sometimes leads to no significant effect at all (Fischhoff, 2012; Kadane & Koehler, 2018). However, Wallsten and Budescu (1995) emphasize that verbal expression may still be valuable, as they "…compensate by conveying greater nuances of meaning" (p.44). One can use rich language to convey the approximate location and degree of imprecision, but at the same time contextualize the uncertainty using "other aspects of the communicator's knowledge or opinions beyond degrees of uncertainty" (p.44). One study found that communicators of numeric uncertainty sometimes want to add additional qualification of uncertainty verbally to serve as “cover” for themselves, especially when they may be responsible for outcomes or their reputations are at stake (Erev & Cohen, 1990).

With respect to forensic evidence used in criminal trials, Martire et al. (2014) found when the underlying evidence was relatively in weak, verbal messages about uncertainty led decision-makers to conclusions displaying “boomerang effects” or “weak evidence effects,” whereby they interpret the evidence as having the opposite valence from that intended by the statements (Martire et al., 2014). If the evidence was strong, they found presentation formats to matter less. In the context of weather forecasts, scholars have made great strides in advancing insights on how to express uncertainty in ways that are “effective” for decision-making. A study by Joslyn and LeClerc (2013) is especially useful. The authors found that people engaged more effectively with expressions of uncertainty when it is catered in two ways: 1) the uncertainty is directly related to the “decision task” in question, and 2) the framing of the message intercepts the tendency for people to choose a “preferred deterministic interpretation” when the cognitive load is too high (p.308). In related work, they found that the condition that effectively counters a “preferred deterministic interpretation” is one that explicitly blocks a common misinterpretation (in their case, this was achieved with statements such as “no rain”) (Joslyn et al., 2009). They conclude that by blocking the misinterpretation, effortful thinking is more easily activated, “allow[ing] for the more complex, probabilistic interpretation and the ensuing benefits to user decisions” (Joslyn & LeClerc, 2013, p. 313).

Beyond messages aiming for effective expression of uncertainty, scientific discourse is full of “hedges” – what Crismore and Kopple (1988) define as a linguistic element that signals tentativeness or caution while expressing information. Journalists, much like human rights practitioners, often consider including those hedges when reporting on scientific findings. Jensen

(2008) tested the effect of including hedging in science news and, found that hedges have little effect on what readers understand about the broader findings presented, but they do impact measures of trust in journalists and of scientists responsible for the research. This is consistent with the theoretical explanation of hedges, or caveats, serving as marks of source credibility.

2.7 Open Questions

In this dissertation, rather than approaching the problem of human rights statistics from the viewpoint of stark good/bad and use/misuse dichotomies, I recognize that numbers are powerful communication tools increasingly in demand in the human rights community. By distinguishing the subset of issues related to the inherent and unsolvable problem of data uncertainty, I seek to understand how human rights practitioners contend with it as they use numbers to inform and influence audiences. At present, we do not have a good sense of how human rights practitioners manage (or not) tensions with such data uncertainty. While there are calls for more transparency about the data generation processes and limitations as one way to contend with data uncertainty, this literature also makes clear there may be at least two potential tensions with this recommendation: 1) Numbers are powerful specifically for their simplified form that do not alert readers to messy, social, and subjective components of their production. 2) More broadly, when science is used in politicized contexts, methodological transparency is associated with reduced persuasiveness and political risk to the information provider. In this dissertation, I aim to understand how human rights practitioners perceive the strengths and risks that may come with being transparent about data they present to their audiences. Do calls for increased transparency have promise as an enhanced communication strategy for a community operating in politically complex contexts?

At the same time, the burgeoning literature advancing more productive engagement with uncertainty in other domains remains inconclusive about effective ways to convey uncertainty to information consumers. However, I draw on the scattered theoretical and empirical insights reviewed here to develop testable hypotheses about the impact of various ways of expressing data uncertainty. Through this study of data and uncertainty in the human rights context, I hope to not only make practical contributions to inform practitioners in this domain, but to contribute to the scholarly literature on communicating data uncertainty.

Chapter 3 Research Methods

3.1 Introduction to Mixed Methods Design

As stated above, this dissertation research sought to answer two overarching questions:

1. How do human rights practitioners think about the value and impact of quantitative evidence and its associated uncertainty when communicating it to inform and influence audiences?
2. When presented with quantitative evidence about violence, crime or abuse, how do different presentations of data uncertainty affect decision outcomes?

To address these questions, I used a mixed methods approach. According to Creswell and Clark (2011), it is useful to employ mixed methods research to enhance the ability to answer multidimensional research questions and gain insights with one method where another method is relatively weak. In my study, I used the grounded theory qualitative research approach in parallel to conducting an online survey experiment (Charmaz, 2006). In general, the qualitative study aimed to answer the first question above, and the experiment was designed to answer the second question. While the two studies are distinct, they are complementary. The insights learned from each one benefited the other, as will become clear in Chapter 6 Discussion.

The strength of the qualitative research is that based on the views, perceptions, and experiences of human rights advocates interviewed, I learned how their context and subjective experiences shape their expectations and thinking about data, uncertainty, and their associated communication practices. The strength of an experiment is that by using randomization as an instrumental variable, one can isolate and identify the effect of a treatment on an outcome within a sample (List, Sadoff, & Wagner, 2011).

The qualitative research informed the survey instrument design. While I drafted vignettes for the survey experiment a priori, I drew on insights with interviewees to improve and adjust the vignettes and interventions accordingly. For example, after hearing time and again from interviewees that they tend to scrutinize information and methodology more when the source of the information was less familiar to them, I modified the vignettes to ensure that the names of the information providers did not resemble any of the more familiar human rights organizations, hoping this would prompt people to pay more attention to the variables of interests in my study. Having these two complementary streams of insight enabled me to conduct what I believe to be a richer design, analysis, and interpretation of the science and politics of communicating human rights data.

3.2 Qualitative Research

To explore how human rights practitioners think about 1) quantification and data, and 2) the various dimensions related to communicating data uncertainty, I began this study with a qualitative grounded theory methodology. Grounded theory enables one to analyze empirical observations to help understand and explain a practice among a community of interest (Charmaz, 2006). While the human rights literature offers evidence that data is influential and communication surrounding quantitative evidence of human rights violations is challenging, little scholarship existed from within the community about how practitioners perceive and navigate these tensions in their daily work. Drawing on Creswell (2013), my qualitative research approach was to use semi-structured interviews to explore multiple dimensions and subjective experiences of participants in order gain evidence about their perspectives. I designed an interview protocol on the basis of my research question, which I piloted before beginning live interviews in Colombia and mostly remote video interviews for participants in the U.S. and Europe (see Appendix A). The semi-structured interviews began with general questions about the organizations work and their communication strategies about evidence and then moved to more specific questions about data production and consumption, quantitatively-inclined terminology, and how they see and communicate about data uncertainty.

For the interviews I was able to conduct in person, I was able to observe participants on their own terms and in the contexts within which they usually operate. In reporting my findings, I aim to be transparent about my own biases and values as a researcher as well as those I detect of the participants. The themes I raise as my qualitative findings were derived during the course of the research and analysis rather than relying on rigid, predetermined definitions.

3.2.1 Population of Study

I conducted semi-structured in-depth interviews with 28 human rights practitioners at 15 human rights organizations who use data and statistics to advance their agendas. This is a very broad, diverse and encompassing group of actors that operate in transnational networks, from high-level international institutions working on a broad set of human rights issues, to very local grass roots organizations working in Colombia engaged in one issue of local importance. I limited my study to individuals at organizations that produce public reports that include quantitative information. As such my sample may be biased toward members of relatively well-resourced organizations, as they likely have the capacity to analyze and publish evidence-based claims to some degree.

These interviewees were from 8 internationally-focused organizations, 2 U.S. focused organizations, and 7 Colombian organizations working to document, expose and advocate for an end to human rights violations. I recruited participants via snowball sampling, drawing on my professional connections to get early interviews and then asking those advocates for introductions to colleagues or to representatives in other organizations. Most of the international and U.S.-based interviews were conducted via video-conference, and all of the Colombian interviews were conducted in person. I agreed to anonymize individual interviewees. All interviewees are referred to with a label (C: Colombian, U: U.S.-based, G: Global) and a unique number from 1 through 28.

Table 1: Interviewees

Organization	Geographic focus	Main issue focus	Primary goal with fact-finding	Number of people interviewed	Work with quantified data
Human Rights Watch	G	HR	Advocacy	8	Rarely produce, often present
Amnesty International	G	HR	Advocacy	5	Sometimes produce, often present

Human Rights First	G	HR	Advocacy	2	Do not produce, sometimes present
Human Rights Data Analysis Group	G	HR	Statistical data analysis	1	Analyze and present
Walk Free Foundation	G	Human trafficking	Advocacy	1	Produce and present
International Campaign to Ban Landmines and Cluster Munition Monitor	G	Landmines	Monitoring & advocacy	1	Produce and present
Southern Poverty Law Center	U.S.	Hate	Monitoring & advocacy	1	Produce and present
Urban Institute	U.S.	Well-being	Policy-relevant research	1	Analyze and present
National Center for Historical Memory (CNMH)	C	HRVs	Historical memory	1	Produce and present
Casa de la Mujer	C	Women's rights	Advocacy	1	Produce and present
SISMA Mujer	C	Women's rights	Monitoring & advocacy	1	Produce and present
Somos Defensores	C	HRD	Monitoring, advocacy & memory	1	Produce and present
Comisión Colombiana de Juristas	C	HR	Advocacy & litigation	2	Produce and present

Instituto de Estudios Políticos y Relaciones Internacionales	C	Social problems	Policy-relevant research	1	Produce and present
UN High Commissioner for Human Rights Colombia Office	C	HRs	Public policy & advocacy	1	Produce and present

My sample from global organizations was skewed towards staff at Human Rights Watch (HRW) and Amnesty International (AI) in the U.S., arguably the two leading human rights organization with a global reach. Their work has influenced the shaping of the human rights movement globally (Keck & Sikkink, 1998; Moyn, 2010). These two organizations are especially relevant for this study because they are large organizations, with many thematic and geographic divisions that do research, conduct advocacy, participate in elite domestic and international policy and diplomatic forums, write public appeals, publish reports for general consumption, and submit evidence to international criminal courts, among other things. Their reports are also coded by peace and conflict scholars and used for academic analyses of patterns of human rights abuses and human rights reporting patterns (e.g. the Political Terror Scale and Uppsala Conflict Data Program) (Gleditsch, Metternich, & Ruggeri, 2014).

It is useful to note that a single organization can have multiple roles with respect to data and analysis. Some organizations, especially those focused on a single country or issue (e.g., Colombia and U.S. focused organizations within this study), have dedicated staff conducting ongoing monitoring and documentation of HRVs for the purpose of counting. In these cases, staff members often review unstructured information from multiple sources (e.g., testimonies, press accounts, documents), contend with their particularities and contradictions, develop definitions and classification rules to structure and code relevant pieces of information and to handle conflicting information, and then digitally organize their data, usually in spreadsheets and sometimes in more complex databases. All of the Colombian groups in this study are data producers to varying degrees, collecting data on specific types of violations or crimes (e.g., killings of community leaders, forced disappearances, sexual violence). Among the U.S. and globally-focused groups, three of the organizations work on specific human rights issues (hate crimes, human trafficking, and banning of landmines). At the same time, almost all the organizations also reuse data (either open source data or data shared from partners) or they cite existing numbers produced by other organizations.

With respect to statistical data analysis, most interviewees consider themselves and their organization non-experts. One group has recently invested heavily in quantitative expertise, bringing in a full-time statistician to run an ongoing global survey. Several groups have a small minority of in-house methodologists or have contracted with short-term statistical consultants for isolated studies using advanced statistical methods. Only one organization focuses exclusively on conducting advanced statistical analysis. Despite varying levels of quantitative expertise, all organizations represented in this study present quantitative claims in one way or another in their human rights work.

My sample of respondents is limited and may be biased. It is a convenience sample originating from my own professional network and extending into branches of my network's network. To try to capture some diverse perspectives, I recruited and interviewed advocates from international, U.S.-focused, and Colombian based organizations. While my own qualitative findings may not be representative of the "human rights community" as a whole, in Chapter 4 I nevertheless draw out the most consistent themes expressed by interviewees to argue what I perceive to be some latent norms governing the use of uncertain data and numbers, at least among this small, yet diverse group.

By including globally-focused groups, U.S.-focused groups and Colombia-based groups, I aimed to cover different perspectives. Global and U.S. groups cover a wider breadth of international issues, and Colombian groups are mostly focused on domestic human rights, political violence and transitional justice related issues (Tate, 2007). All of the groups operate under the same international human rights framework, sharing some communication strategies and data use practices, but working in different contexts, with different level of resources, and targeting some different agendas and information consumers.

3.2.2 Analysis

I received consent to record all interviews. I then transcribed then, coded them using N-Vivo software, and analyzed them inductively to detect prominent themes across the set. A research assistant and I first read half of the interviews with the sole purpose of building the codebook. I identified naturally occurring ideas in the interviews, clustering them by common themes, and repeating the process until meaningful themes emerged for interpretation (Charmaz, 2006; Creswell, 2013). I then both re-read all English interviews and applied the codes systematically. I then coded all the Spanish interviews with the same codebook. All interviewee quotes are kept anonymous in the findings presented in Chapter 4: Qualitative Interview Findings.

3.3 Pilot Study of Sample of Human Rights Watch Reports

To complement the perspectives gathered through interviews, I also conducted a small pilot study of Human Rights Watch publications to see if I could gauge the extent to which they present statistics in their report, their inclusion of methodological detail and the language they use to convey any limitations to numeric claims. I selected HRW publications because HRW is arguably the leading human rights organization in the world, and because the interview study also included a greater perspective from HRW staff than any other organization, so it made sense to compare with their reports. HRW reports are also a key publication included in many quantitative indicators about HRVs in the world (Cingranelli & Richards, 2010; Fariss, 2014). After compiling a complete list of all HRW reports from 2012 and from 2018, my research assistant and I selected a simple *random* sample of six reports from each of the two years included. I selected these two years somewhat arbitrarily, based on one interviewee informing me that sometime in the mid 2010s HRW enhanced their focus on methodology sections. We coded the reports for five main variables: presence of quantitative claims about human rights violations, the purpose the quantitative claims was serving in the report; presence of data limitations about quantitative claims; the type of limitations message; where in the report the data limitations message was included (e.g., alongside the quantitative claim, in a methods section). We then conducted simple descriptive analysis of the coded variables.

3.4 Survey Experiment

3.4.1 Experimental Design

I designed the survey experiment to answer the following primary research question:

RQ: When presented with quantitative evidence about violence, crime or abuse, how do different ways of conveying data uncertainty affect decision outcomes?

Using an experimental design allowed me to examine the causal effects of different presentation messages and formats of communicating data uncertainty on individual data-informed decisions related to human rights, while controlling for potentially confounding variables.

I developed theoretically-informed hypotheses and treatments (presented below) on the basis of literature reviewed in Chapter 2: Literature and Theoretical Framework. To test my hypotheses, I carried out a between-subject experiment with 6 experimental conditions (2 controls and 4 treatments) which I tested across three different vignettes (6 x 3 design). I recruited 970 individuals via Amazon's Mechanical Turk (MTurk).

3.4.2 Treatments

The language used for each treatment assumed to draw on the data presenters knowledge of how the data is uncertain – in these cases that numbers have high uncertainty and disproportionate selection bias in one group – and tests the effect of making the data uncertainty explicit to decision-makers in various ways, the goal here being to help decision-makers avoid interpreting the numerical information with false precision. For example, in one of the vignettes designed for the study – a hypothetical vignette involving rape survivors – the data producers know, on the basis of working in the local context, that underreporting is probably much more severe in one region (Choco) than in another (displaced persons camp in Sucre). The treatment messages draw on the data provider's insight about the provenance of the data and convey the data's likely limitations. In this case, an "effective" adjustment would be to allocate a relatively higher proportion of the available resources to Choco to account for the possibility that in Choco, there may be considerably more victims that could require services than those suggested by the reported number of victims (as compared to the control condition when raw numbers and no uncertainty are provided).

Each participant was randomly assigned to an experimental group (among four treatments and two controls). Respondents then received the same treatment for three vignettes, which they viewed in random order. For each vignette, after asking respondents to make a specific decision (all on continuous scales), respondents were asked to rate their level of trust in the institution providing the information on a Likert scale. In addition to these main outcome measures, I also collected data on respondents' confidence in their decision and their self-assessed level of familiarity, and expertise on the specific issues covered in the given scenario. After viewing the three vignettes and related questions, the survey concluded with questions about respondents' political views, math education, quantitative literacy, cognitive reflection, and personal demographics, all to serve as controls in the analysis.

Each participant was randomly assigned to an experimental group (among four treatments and two controls). Respondents then received the same treatment for three vignettes, which they viewed in random order. For each vignette, after asking respondents to make a

specific decision (all on continuous scales), respondents were asked to rate their level of trust in the institution providing the information on a Likert scale. In addition to these main outcome measures, I also collected data on respondents' confidence in their decision and their self-assessed level of familiarity and expertise on the specific issues covered in the given scenario. After viewing the three vignettes and related questions, the survey concluded with questions about respondents' political views, math education, quantitative literacy, cognitive reflection, and personal demographics, all to serve as controls in the analysis. Before launching my full study, I conducted a pilot study to test for any unforeseen issues with my instructions, instruments, data collection software, and to get a better sense of the time it would take respondents to complete the full survey.

3.4.3 Vignettes

I included three vignettes (as opposed to one) to control for any vignette-specific effects. Lab findings are more robust if treatment effects hold across the three decision-making scenarios. Across vignettes, I assume that information providers are local experts, thus in the position to know that the available numbers they present are partial and biased due to limitations in reporting and observability. Each vignette was designed to do the following: 1) present a scenario that reflected, as closely as possible, a real world human rights situation and realistic decision context; 2) use data about violence or crime collected for some other original purpose and reuse it in scenarios requiring a decision; 3) require the decision-maker to put themselves in the shoes of a professional decision-maker; 4) make a resource allocation (in two vignettes) or criminal sentencing decision (in one vignette) where the numeric information could have a consequential anchoring effect. Overall, the scenarios were such that misinterpreting the numeric data provided, for example by seeing a trend where one may not actually exist or drawing a seeming clear but incorrect conclusion, could be problematic. I wrote the vignettes to try to make the information and uncertainty messages the resounding signal, minimizing other signals people often heuristically use in decision-making with statistics and uncertainty. For example, name recognition and professionalism are often used as heuristics for judging the value and credibility of claims (which advocates discuss as a kind of heuristic in the interview study). To try to weaken the potential role this might play, even in fictional scenarios, I tried to avoid an institutional name that might sound overly professional. I aimed to take an objective and measured tone, using informational framing, and avoiding emotional or motivational framings, so as not to introduce confounds from other messaging appeals (McEntire, Leiby, & Krain, 2015a). For example, I did not include images and colors, as they have been shown to affect how individuals respond to materials, and in particular, to potentially elicit different emotional reactions (Small & Verrochi, 2009; Valdez & Mehrabian, 1994).

Table 2: Overview of Vignettes

Vignette Issue		Vignette format	Numeric Information	Decision task	Related Literature
	Health Kits for Wartime Sexual Violence	Humanitarian organization asking for input on how to allocate limited health resources to rape survivors across two locations.	Number of rape survivors documented by the organization in two regions. Large, uncertain region: 200 documented cases Small region with higher surveillance: 400 documented cases.	Allocate a portion of 700 available health kits to the region with most uncertainty in the reported data.	(Cohen & Green, 2012; Leiby, 2012; Merry, 2016; Roth et al., 2011)
	Prosecution Resources of War Crimes Trials	Assistant to the Office of the Prosecutor at a Special War Crimes Court must advise on how to prioritize allocating limited investigative resources to cases along ethnic lines.	Number of known civilian casualties by ethnic group, as submitted by a human rights group who documented killings during the conflict. Minority group: 3,600 Majority group: 400	Prioritize how many of 80 cases to investigate for alleged perpetrators from two ethnic groups	(Reydams, 2016)
	Evidence-based sentencing	Sentencing judge must determine the sentence length for a man convicted of a felony based on summary of the case.	Recidivism risk score based on available crime data calculated by criminologists & provided in probation report. Risk score: 8 on a 10-point scale (equivalent to "high risk").	Determine number of years (between 0 and 20) to sentence the convict to prison.	(Hannah-Moffat, 2013; Starr, 2014)

Each vignette presents a scenario that reflects, as closely as possible, real world human rights situations and realistic decision contexts. Vignette 1 presents convenience data about known cases of sexual violence. Reporting about sexual violence, and in particularly wartime sexual violence, is notoriously incomplete (Cohen & Green, 2012; Leiby, 2012; Merry, 2016; Roth et al., 2011). Preventing and supporting victims of sexual violence is also a priority issue for many human right grant-making organizations. Thus, Vignette 1 presents a typical scenario where grant-makers must allocate resources among two contexts with highly uncertain information. Vignette 2 is inspired by political science professor Luc Reydam's 2016 study on the influence of an NGO on the International Criminal Tribunal for Rwanda (ICTR). He argues that the selection bias in a large body of evidence presented by an NGO in the early days of the ICTR may have led to bias in the prosecutorial strategy, which he argues in the long-term may be part of the explanation for the ongoing impunity against the Tutsi perpetrators of violence during the 1994 Rwandan ethnic conflict. Thus, Vignette 2 creates a fictitious, but hopefully believable situation where criminal court investigators must allocate prosecutorial resources on the basis of partial collections of evidence. Given that Starr (2014) found evidence that actuarial risk scores had a significant impact on sentencing outcomes among law student decision-makers, and given that she interprets the veneer of objectivity as likely influencing this effect, I have included here a version of the evidence-based sentencing (EBS) scenario she designed as Vignette 3 to see if adding data uncertainty interventions create any attenuating effects on her observed effects. The use of risks scores and other predictive measures in the criminal justice system are increasingly an issue of concern in the human rights community (Lum, 2017; Lum & Isaac, 2016).

3.4.4 Treatments

I designed four uncertainty message types (treatments), drawing on theoretical insights from the science communication and practical insights from human rights scholarship and practice.

T1 – Simple Caveat

Recall that theoretically, caveats signal caution while expressing information and are often symbolic over substantive (Crismore & Kopple, 1988). In studies about the presentation of science in the media, caveats have been found to signal credibility if communicated by authors of claims (Jensen, 2008) or to have no detrimental effect on perceptions of source trust (Retzbach & Maier, 2015). In human rights reports, practitioners either use simple caveats to communicate research limitations or "approximate" language to warn of potential under-reporting (Satterthwaite and Simone 2016; Heinzelman & Meier, 2013). I tested the impact of this message type, collecting information both on decision outcomes as well as perceptions of trust of the information provider.

T2 – Rich Background

One of goals of transparently communicating data limitations is to put on record much more information about earlier stages in the knowledge production process. Responsible data scholars are recommending that when data is used to make claims about magnitude, patterns, and related claims, human rights communicators should also provide information about what is known and knowable – i.e., the data generating process – and disclose the known data flaws more precisely and fully than what is done in the typical caveat. At the very least, such descriptive information can be useful as documentation for data reuse and posterity.

However, the literature on communicating uncertainty finds outcomes in the presence of long and non-numerical disclosure of data provenance as potentially unfavorable. More description about data provenance could be interpreted with great variation by audiences, it could generate cognitive overload (Kahneman, 2011), or it could be ignored altogether (Lupia, 2013). The effect on decisions or source credibility of providing more information about data provenance remains unclear, with conflicting claims in the literature. I tested language that goes beyond the caveat, adding precision and detail about the data production and limitations for all data presented.

T3 –Expert Interception

One of the most useful insights about communicating uncertainty came from Joslyn & LeClerc (2013) that found people significantly improved their cognitive engagement with uncertainty information, and made more cautious decisions, once a "preferred deterministic decision" was intercepted with language directed at the decision task. In human rights domains, scholars' greatest concern about the use of uncertain numbers is that audiences receiving the information will confuse data patterns as real-world patterns and that such erroneous inference and interpretation will lead to bad outcomes (although what constitutes good and bad outcomes is decision-specific and context-dependent). The question remains: if we think such misinterpretation is problematic, and we think that more effortful thinking about data uncertainty is beneficial, what would be the impact of directly foreseeing and intercepting this erroneous interpretation? Would doing so lead decision-makers to alter their decisions? I designed and tested language conceptually guided by Joslyn & LeClerc's insight, explicitly drawing on the expertise of the communicators to warn about the potential bias in data and the impact of that bias on the decision task.

T4 - Numerical Expression of Uncertainty – Confidence Intervals

As seen above, numeric expressions of uncertainty are often more reliably interpreted than linguistic expressions, as they more precisely express the direction and degree of uncertainty. At the same time, we know that confidence intervals are often ignored (Tversky & Kahneman, 1974). I included a treatment where uncertainty is expressed as a numeric range to indicate the skewed direction and degree of uncertain numbers. The treatment aims to numerically express what is verbally expressed to varying degrees in other treatments – the sizable selection bias in data underlying numeric claims. While these may be more reliably interpreted, we do not know the effect on decision-making in their presence, especially because there could be the confounding effect of heuristic reasoning with statistics.

If we find that numerical expressions improve decisions reliably, it could be a case for further investment in statistical methodologies for these applications and for producing formal estimates in more human rights projects. On the contrary, if the outcomes with this treatment are equal or relatively less reliable than verbal forms, it could guide advocates to concentrate on linguistic forms of communicating data uncertainty, as tested in other treatments.

3.4.5 Controls

I included two control conditions:

C0 - No numbers

A pure control condition, with no numbers included to influence the decision-maker. This condition was included as a way to gauge the population "prior" – how do people make decisions in the absence of numbers.

C1 – Numbers, No Uncertainty

A modified control, numbers presented as objective and hard evidence of the magnitude and patterns of crime or victims of violence. The idea here is that numbers may be seen as hard evidence, with the veneer of precision and objectivity intact.

On the basis of the theoretical insights reviewed above, I propose the following hypotheses about the impact these treatments on two primary dependent variables: decisions and trust perceptions of the information provider:

- H1: Quantitative information presented with its full veneer of authority and objectivity – and with no accompanying information to disclose data uncertainty – will shift the mean of decision outcomes relative to when no quantitative information is presented at all (C0 is different from C1.)
- H2: Simple Caveat – being short statements signaling tentativeness or caution with data – will have no effect on decision means relative to presenting information with its full veneer of precision and objectivity. This would suggest evidence of the "caveat fallacy." (C1 is equal to T1)
- H3: Expert interception on anchoring effect in reported data (T3) will lead to shifts in decisions relative to simply providing audiences with more information about the data generation process (T2).
- H4: Given that uncertainty statements are ascribed to the information provider, trust in the information providers will benefit from disclosing data uncertainty (i.e. there will be higher levels of trust in the information providers when some expression of uncertainty in data is provided Treatments > C1).

Table 3 summarizes the hypotheses.

Table 3: Primary Hypotheses

Treatment	Type	Characteristics	Main Decision Hypotheses
C0	No Numbers	Minimal information	
C1	Number + No Uncertainty Qualifiers	Numeric	C1 ≠ C0
T1	Number + **Simple Caveat**	Language, short warning that uncertainty exists	T1 = C1
T2	Number + **Rich Background**	Language, long description describing possible sources of uncertainty stemming from data generation process	?
T3	Number + **Expert Interception**	Language, long description warning of uncertainty and corrective decision-specific suggestion	T3 > C1
T4	Number + **Numeric Confidence Interval**	Numeric, short description of a confidence interval	?

I do not feel the current state of theory enables one to make strong hypotheses on the effects of other treatments. For this reason, beyond testing these hypotheses, the study will also enable exploratory observations to guide future research.

The post-vignette survey questions will include questions on participants demographics (i.e. age, gender), level of math education, political orientation (based on political typology questions) and cognitive reflection (Frederick, 2005).

3.4.6 Example Vignette and Treatment Language

I include here the full implementation of one vignette to guide reader as they consider the design and view the results. The other vignettes and the specific language used for the controls and treatments are included in Appendix B.

Vignette 1: "Health Kits" for Wartime Sexual Violence

Stop the Violence Now (SVN) is a non-profit agency dedicated to promoting human rights globally. We are headquartered in New York, but collaborate with local authorities around the world to eradicate violence and its effects.

We are concerned about the long-term health of rape survivors. Studies show that survivors of sexual violence are more likely to experience mental health issues such as depression, anxiety, substance abuse, and post-traumatic stress disorder. They also have higher rates of illnesses like cancer, cardio-vascular disease, and diabetes over their lifetimes.

We have decided to provide post-rape care resources to victims of sexual violence in war-torn countries. These come in the form of "kits" – which include a bundle of professional services promoting physical and mental health, and social services.

We will begin our provision of these kits in Colombia. This year, SVN's offices documented increasing numbers of sexual violence in two distinct locations in the country. One is the large and remote Pacific region of Choco where there is still an armed conflict. The other is a displaced persons camp in the northern region of Sucre. We will distribute emergency health kits in these two regions.

[* C1+ Treatment]

We aim to distribute our limited services to benefit as many victims of sexual violence as possible who arrive at the locations where we offer services. Often, we find that more people show up for services than official counts have tallied.

We currently have the resources to fund 700 kits. We are seeking input on how to best distribute them.

Based on the information provided here, please tell us how many of the 700 kits you would allocate to SVN's office in Choco: _______

Treatment	Name	Uncertainty message [*]
C0	No numbers	[No C1 paragraph above]
C1	Number, NO uncertainty	In Choco we have 200 documented cases of sexual violence and in the Sucre displaced persons camp we have 400 documented cases.
C1+ T1	Simple Caveat	Be aware that reporting sexual violence is culturally stigmatized in Colombia. Many people may choose not to report their experiences for fear of social consequences.
C1+ T2	Rich Background	In both regions of Colombia, people may avoid reporting sexual violence due to cultural stigma. In Choco, local government-run health clinics are doing their best to report known cases, yet there are very few of these clinics across the large region. More generally, data collection is challenging in a context where illegal armed groups continue to operate and use sexual violence as one way to repress the population. In Sucre, all the reported cases are from a relatively small displaced persons camp. There are several health clinics throughout the camp where people can get aid, and local officials make an effort to document all the sexual violence cases they learn about.
C1+ T3	Expert Interception	Be aware that the reported cases of sexual violence are an undercount, and likely do not represent the total number nor the pattern of all sexual violence cases in both locations. While we cannot be sure, we may know about many more of the victims of sexual violence in the Sucre camp than we do about all victims in Choco. In Choco, there are probably many more victims of sexual violence than are reported because the region is so large, precarious, and still has active armed groups. In Sucre, there may also be more victims than the reported number, yet the displaced persons camp is smaller, has administrators working around the clock, and is not located in an active conflict zone.
C1+ T4	Numeric uncertainty, conference interval	[C1 embedded] In Choco we have 200 documented cases of sexual violence, but the total number could range to be anywhere between these 200 to 2,000 cases. In the Sucre displaced persons camp we have 400 documented cases, but the number of people who have suffered some form of sexual violence could be as high as 600.

3.4.7 Subject Pool and Sample Size

As mentioned above, I recruited 970 college graduates via MTurk based on a conservative target sample size of 1,000. A priori, I calculated that I would need at minimum a sample size of 678 to achieve 80% power based on an assumed effect size of f=0.15 (computed using g*power (Faul, Erdfelder, Lang, & Buchner, 2007)). To determine the assumed effect size, I drew on values reported in Starr (2014), who ran a lab experiment with a similar structure in the context of criminal sentencing with risk scores. The study reports an 0.8 year increase in sentencing length in the high risk + risk-score condition (p. 869). I calculated the initial effect size assuming my control group would have the expected mean in Starr's high risk + risk score condition (5 +.8 years). I then made the remaining assumptions relative to the 5.8 years in the control condition. These assumptions were highly conservative given that in some conditions, the literature reviewed did not given me a strong basis to develop clear hypotheses. Then, I calculated an SD α of 2.14 years by calculating Starr's residual standard deviation based on her reported standard errors. For my calculations, I assumed equal residual variance across my treatment groups. I included a Bonferroni correction to account for analysis of two dependent variables, yielding a significance level of .025 based on a target family-wise error rate of 5%.

Studies have repeatedly shown that MTurk generates reliable, high-quality data for experimental social science research (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). Mturkers were eligible to respond to this the survey if they met the following conditions: they were based in the U.S., had at least a college degree, and had a successful track record of completing tasks on MTurk (over 1,000 tasks previously completed with a 95% approval rate). On MTurk, the job ad avoided mentioning human rights, politics, or any of the issues covered in the vignettes. It simply asked participants to provide their best judgment in relation to the scenarios. Respondents were paid \$1.50 for responding to the full survey, calibrating to other vignette-based work on MTurk (Berinsky et al., 2012; McEntire, Leiby, & Krain, 2015b; Paolacci et al., 2010). Buhrmester et al. (2011) found compensation to only affect the number of participants who complete the survey, not the quality of their responses.

Chapter 4 Qualitative Interview Findings

In 2013, Colombia's National Center for Historical Memory (CNMH) released its flagship report, *Basta Ya! Colombia: Memories of War and Dignity* (2013), recounting over 50 years of violence in the country's armed conflict.[1] One of the report's most visible and memorable findings was that at least 220,000 people died as a result of the decades-long Colombian conflict. In addition, its leading chapter reports many other prominent statistics, including the proportional responsibility of conflict perpetrators for massacres, trends in forced disappearances, kidnappings, sexual violence, and forced recruitment across 30+ years. The report has been heralded by the press, civil society, and even critics as being an extremely valuable contribution to historical reckoning.

When I spoke with one of the report's main authors in November of 2017, he confessed that the 220,000 figure and the report's overall reception far-exceeded his expectations in press coverage, impact with the government, and favorable public reception.[2] At the time of publication, he worried about what he felt were severe limitations in the data underlying the reported statistics. CNMH has had to collate cases from many different sources with varying data quality and opacity in the underlying records, making the merging of duplicate reports difficult. Some of the available streams of data, he suspected, likely had biases in their reporting practices. Documentation of some types of violence were only recently becoming possible, such as

---

[1] "*Basta Ya! translates to "Enough Already!*."
[2] Participant C8.

evidence of paramilitary crimes.[3] "When we launched the figure we had great fears, because there were methodological gaps that could be used to undermine us." Yet, he felt real pressure to include statistics: "we could not come out with such a report without putting on the public stage the issue of the sheer dimensions [of the violence]," which the numbers helped portray.[4] To manage the tension, the team decided to lead with the 220,000 figure which they felt was a lower-bound based on their data cleaning of reported killings across multiple sources, but they also warned readers "not to confuse data with facts."[5] The first section of the volume elaborates on the difficulty of documenting and measuring the experience of violence. Accompanying the report was also a three-minute video describing some of the limitations of the CNMH's violence databases.[6]

While CNMH staff braced themselves for critique, criticism of the limitations of the statistics never came. "We presented them, but the impact of the 220,000 was so strong, for example, in the media, that no one asked about the rigor used to construct that statistic or how we arrived at it." There was little appetite for nuance by the media, which he believes is because "the magnitude of the figure was so impactful on the public stage," but also because of the level of trust the institution had established within Colombian society. "When one builds a reputation…society doesn't ask much about the route….depending on the audience, what matters

[3] In 2003-2006, the largest paramilitary organization demobilized and agreed to confess crimes to be eligible for reduced penalties.
[4] In fact, he says, the CNMH had shied away from publishing statistics in earlier reports. The CNMH's experience reporting the phenomenon of forced disappearances is telling. Due to the difficulty of comprehensively documenting this type of violence, the CNMH's first report on forced disappearances had opted for not leading with statistics. It instead had a whole chapter talking about the limitations of knowing about forced disappearances, given that perpetrators intention to erase the person, the event, and the possibility of knowing altogether (CNMH 2014). Yet, they were surprised by a strong call for a magnitude estimate, thus the CNMH wrote a second report, in which they report 60,630 victims between 1970 and 2015 (CNMH 2016).
[5] He also shared how they considered yet discarded the option of calculating statistical inferences rather than reporting numbers based solely on documented cases, but he says he was advised against it. International advisors stressed that numbers needed to be based on names, as statistical estimates in other countries' had been perceived as elusive, with a public that was not accustomed to thinking formally about statistics and uncertainty in events as concrete as death.
[6] http://www.centrodememoriahistorica.gov.co/micrositios/informeGeneral/basesDatos.html?fbclid=IwAR2X6hRmP8gLlnZOP0ied1TaowkVTQrHAbTyrFqN9uHOYiI4EiuVCqM2YxQ

is the figure and the producer."[7] The organization did not even get pushback from political actors potentially threatened by the claims. He was most surprised that some sophisticated academic peers and methodologists, people he felt would have clearly seen the deficiencies in the data and analysis, did not publicly raise concerns.

A few days after my interview with him, I spoke with one of the academics that the CNMH's representative viewed as a potential leading critic. I asked this scholar about the report's statistics. Without hesitation, he described them as "dreadful," but then, he went on to consider the overall value of the report nonetheless:

> "it's a report that talks about the horrors of the war, that says that the state was complicit but at the same time doesn't side with the guerrillas, that in the Colombian context is something new…it obliged [President] Santos to apologize to the country, so it's a report that caused such political sympathy for reasons that are so good, that one does not want to get in the way of that…"[8]

Overall, his assessment was that it was unwise to quibble over the data and methods in the face of such an important success for historical reckoning in the country. He explained his reasoning:

"So there is a criteria that is academic and one that is political, so with[in] which school of thought shall I stay? So let's say that I try to stay within a bound that is tolerably bad…Is it worth removing [the statistic] or not? I'm not too sure, this is probably an ambiguous response."[9]

This exchange captures much of the tension the human rights advocates I interviewed deal with as they make quantitative claims on the basis of uncertain data, both in Colombia and beyond. They struggle with a tension between the demand and supply of data, and the politics

[7] Consistent with the idea of strong trust in sources perceived as experts (Lupia 2013; Andreas and Greenhill 2010).
[8] Participant C7
[9] Participant C7

and practicalities of evidence-based claims-making in favor of human rights advocacy. The numbers, even when uncertain, can be of great value when they generate a social and political response, but they must steer clear of "bad statistics." In this chapter, I analyze and find a relatively consistent and stable set of themes which I will argue constitute informal norms regulating "good enough" numbers for human rights advocacy: imperfect, temporary, conservative and empirical, i.e., they are based on available data rather than unfounded guesses. "Good enough" numbers are not intended to be precise – and may indeed be far off from representative claims – but by and large, both producers and their audiences find them acceptable for the primary goal of garnering attention to human rights violations. The role of transparent communication of data production and data uncertainty has an ambiguous role in this tenuous formula. One the one hand, transparency is seen as a generically good, responsible practice by most respondents in this study, yet many emphasize that audiences have little interest in it unless the numbers are glaringly questionable or if the source provider is weak.[10]

While the CNMH researcher remains self-critical and self-conscious of the reported figures, he finds solace in the fact that historical clarification and memory work continue – emphasizing a temporary quality to the numbers in the CNMH's reports. Among their ongoing efforts, they established more permanent monitoring through their violence "observatory" so that they could take more direct responsibility over data collection going forward. They found that since *Enough Already!* was published, many more people contacted the CNMH to offer new testimonies. By the time we spoke in 2017, the number of documented conflict-related deaths

[10] In the context of this study, weak usually meant the organization was not well-known or had some previous history of being questionable.

had grown to 275,000. The original report, he told me, was never meant to be the final word on conflict-related statistics. The construction of historical memory is participatory and ongoing.

Drawing on interviews with staff from leading human rights organizations in Colombia, in the U.S., and at the international level, this chapter offers the perspectives of human rights advocates: how they 1) view their work producing and citing quantitative data and statistics about human rights abuses and 2) perceive their communication strategies about numbers and data limitations to their respective audiences. As the exchange about the CNMH's war-related death statistic highlights, most of the study participants use the rhetorical power of numbers to advance their advocacy goals. These perspectives are additionally valuable because they offer insight into how advocates *intend* for numbers to be read. Having this viewpoint allows us to later compare the intentions of human rights advocates with the experimental trial that, in part, offers insight into how numbers are actually comprehended and used by laypersons as input to decision-making.

Satterthwaite and Simone (2016) encourage a move away from contrasting human rights research with social scientific standards, to a study on existing and possibly shared "norms and principles" regarding the use of quantification in human rights advocacy. In this vein, one of the contributions of this chapter is to present what advocates (in this study) quite consistently voice as their notions of "good enough" numbers for advocacy. I analyze what appears to be the guiding and constraining principles of "good enough" numbers, how interviewees determine what counts in this category, and why such numbers appear to be generally accepted in the human rights community without rebuke. While "transparency" is often stated to be a core ingredient of "good enough" numbers, in practice, advocates say they do relatively little, often less than they would like, to convey the methodological detail, information about the data production process, or any limitations they know to inherently affect their quantitative claims. Ultimately, conveying data uncertainty is secondary to more pragmatic choices advocates make as they use quantitative information to advance their advocacy goals.

4.1 Advocates emphasize value of numbers in advocacy

Human rights advocacy is essentially a persuasive endeavor. All 28 human rights advocates interviewed for this study – whether working on hate crime issue, sexual violence, lethal violence, or any other human rights violations, whether in Colombia or elsewhere – confirm that quantitative, data-based claims are highly valuable to compel favorable attention to their causes among multiple audiences (confirming what other human rights scholars have noted). Primary audiences for numbers are journalists (and "the media"), policymakers, and the general public. Based on my interviews, numbers serve the advocacy agenda in three main ways: 1) numbers communicate crisis and thus compel highly-desired attention; 2) simple descriptive analyses of data bolster primarily qualitative claims; and 3) to a much more limited degree, advanced statistical analysis can bring scientific "cover" in complex politicized contexts. Despite the three implying different degrees of quantitative expertise and statistical complexity, interviewees stress that all three are inherently used for their *rhetorical value*. For this reason, many interviewees describe recent shifts to incorporate more data and statistics into their work.

4.1.1 Advocates find that numbers compel attention

In my interviews, I find that by far, the main way advocates view the power of numbers is in relation to the media. However, they describe the numbers serving less as precise counts and more as triggers to draw much needed *initial* attention to issues. I emphasize initial because numbers, and press attention more generally, are meant to be snapshots, with little concern or consequence if they become outdated quickly. In the words of one researcher, “The value is in that everybody wants to know [a number]... It attracts attention. It attracts media attention, it...can be really effective and have expedience.”[11] Another high-level advocate echoes the sentiment, “…having been a journalist before, I remember how important it is to put numbers on a certain crisis.”[12] At a basic level, by providing a number about the magnitude of a crisis, they not only trying to heighten attention to an issue, they also believe they are answering what audiences have come to expect in order to remain engaged. In this way, numbers may serve as effective stimuli to capture information consumers’ very limited capacity to pay attention to new information (Lupia, 2013).

All interviewees suggest, in one way or another, that while numbers are quantitative units, they actually used to communicate a *qualitative* message about whether the issue in question is devastating enough to deserve limited attention in a saturated attention economy. “It's more about the ‘what's happening’ rather than the ‘how much,’” said another participant.[13] So, while it is a specific number that is provided, the message meant to be unambiguously heard by audiences is ‘it’s a lot.’ Another participant put it this way: “It flags for people the extent, and maybe the urgency of, a problem, like attacks on human rights defenders around the world.”[14]

[11] Participant G19
[12] Participant G26
[13] Participant G23
[14] Participant G10

One interviewee in Colombia emphasized, “in fact, if you do not use that language, seldom would you make the headlines, I would think, because it becomes a situation that automatically turns on an alarm…don’t put adjectives…put the statistic.”[15] Several participants believe that numbers make issues memorable, and that is required for sustaining their audiences attention. For C1, numbers “stay in your memory and they create public opinion and media pressure for an issue.”[16]

Monitoring groups, i.e., groups collecting data about a given phenomenon in an ongoing way, feel that producing data and updated counts has enabled them to make abuses socially and politically legible over time. For example, one U.S. interviewee said, “I can also tell you anecdotally we get tons of press hits off of the hate [group] list. It's the reason it's kind of the biggest thing in our marketing and publications… It's much easier to place a story that way than it is to saying hey, hate groups are a problem, that's a weaker claim or a weaker position to advocate from than the numbers.” Similarly in Colombia, several groups described numbers as a way to ensure journalists maintain attention on an issue over time, because they can provide numerical updates, what one interviewee called a “media coup.”[17] [18]

Beyond the media, participants talked about the value of numbers with policymakers in a similar way, to convey ‘it’s a lot.’ In the words of a high-level advocate, “if we are able to tell you that over 6,000 people have been killed in the Philippines as a result of the drug war being waged by the president, that counts with decision makers. Of course, they always want to know

[15] Participant C3
[16] Participant C1
[17] Participant C2
[18] Participant C1

the stats."[19] Numbers serve "just to impress upon policy makers that this was not a marginal or small issue."[20] To these audiences, conveying numbers is not only meant to make issues visible, but to discourage issues from being "downplayed"[21] or considered "isolated cases."[22] Another interviewee described the pressure she feels from policymakers to put a number on problems that are hard to quantify. "We know the problem is big enough to do something about, and on the advocacy side, we [get asked] those questions all the time... We started saying up front, we don't have the data,…but yeah, I mean, it would be really helpful to have those bigger numbers."[23]

This practice of reporting numbers over time did invite a few lines of cautious critique. One advocate said the allure of the number is sometimes too good: "The plus side is that the number helped us convey the severity of [the problem]. The downside is that the number has now lived in infamy and people continue to cite the 300,000 figure even though it's 20 years old and is no longer anywhere near accurate. It's like once you put a number out into the universe, it's very hard to get people to shift."[24] While anchoring is a core behavioral heuristic (Tversky and Kahneman, 1974), here we can appreciate their staying power decades passed a number's original moment of use. This power – and danger – is exacerbated by the facility with which numbers are moved from their original context of production – "mutant statistics," as Best (2012) calls them. Another concern is the age-old body counts problem, where a metric becomes a target in its own right. There is prominent, yet subdued concern that the allure of numbers in the press especially creates incentives to always want to show things as getting worse. It also leads to creative arithmetic calculations to make the available data more dramatic, like choosing

[19] Participant G23
[20] Participant G22
[21] Participant G26
[22] Participant G20
[23] Participant G17
[24] Participant G22

to report incidents in a more dramatic unit of analysis, such as "every 29 minutes, a Colombian becomes a victim of sexual violence."[25]

In this way, advocates (especially in Colombia) recognize that numbers used in this way are "superficial …very rhetorical, and one sees that the data do not really serve to describe problems or make decisions."[26] We know from Porter (1995) that the force of numbers in public life is that their consumers mechanically accept them as objective, authoritative information sources, especially in contexts where there is relatively weak subjective authority. What we see here, however, is that their use, while guided by some conscientious parameters (as will become more clear below) is not aligned with how audiences tend to understand numbers (as objective and powerful anchoring informational inputs.). We begin to see the mismatch between human rights advocates' performative use of numbers within a broader persuasive strategy with what theory (e.g. Porter, 1995), and empirics (e.g. Kahneman, 2012) demonstrate to be the way audiences cognitively process and use quantitative information. The anchoring effect of uncertain numbers about abuses, violence, and crime is further corroborated in the experimental survey conducted as part of this dissertation, see chapter five. The key insight we gain here is that advocates see themselves as primarily using numbers performatively to a much greater extent than they use them to convey any kind of consequential "truth-claims."

4.1.2 Advocates believe numbers offer complementary, descriptive value

The belief that numbers are quantitative short-hand for qualitative statements meant to convey severity gets a bit tricky as advocates simultaneously describe what they deem as

[25] Participant C3
[26] Participant C3

secondary efforts to use quantification more analytically. Beyond capturing media attention, many interviewees share the conviction that their organization's most important work is qualitative in nature; any quantitative component is only complementary. As G10 said, "We're using quantitative research to bolster the qualitative, case-based research that we're doing on the ground."[27] The approach is usually to try to derive some analytic value from data by conducting simple descriptive analyses of the information available to them. Typically, interviewees describe plotting data patterns or simple statistics like percentages, averages and rates as a way to amplify what they know about a phenomenon in an anecdotal way. "It's really like the stories of the individual person who's affected, coupled with the data, [used to] impact those decision makers so they get on board, or at least get moved to neutral, those kinds of things. Those are really our goals."[28]

This complementary role of quantification is sometimes described by interviewees as a way to conduct "sanity checks" on how they understand the qualitative investigation of the organization. One such interviewee emphasizes that in a qualitative way, "we're gesturing as accurately as we can at a large truth, which is that, I can anecdotally tell you from my experience as an analyst, who the players are, which groups are rising, which groups are falling, which ones are staying about the same."[29] However, advocates' use of numbers as corroboration to their qualitative insights can be a slippery slope into what Best characterizes as non-intentional misuse of numbers. For one, the imperfect nature of data means that data bias could lead to confirmation bias. However, because advocates downplay the overall role of numbers in their research process – and many times quantification work is done by non-experts or temporary consultants as

[27] Participant G10
[28] Participant G21
[29] Participant U28

resources permit – the overall belief by advocates about the inconsequential nature of quantitative claims (as seen in their rhetorical use) conflicts with their attempts to use data more analytically. The positioning of simple analytic quantitative claims as secondary may cloud their recognition of such bias. Viewing numbers as having secondary role also conflicts with what we know to be the strong anchoring power of numbers in information processing by audiences.

Analytically, what advocates say they most want to do with data and numbers is make comparisons to emphasize their points. In an example from Colombia, one organization that provides descriptive analysis of violent incidents emphasizes that their goal is to make claims about escalation, improvements or simply the unwavering continuity of repression. C4 credits his organizations' ongoing statistical monitoring of violations over time and perpetrator to bringing much needed attention to the escalating and permanent violence in Colombia in the 1980s, 90s and 2000s. "Although with work that was not too sophisticated, but was judicious," we were able to reach "not only the Colombian population, the Colombian authorities, but also the members of the UN Human Rights Commission, UN subgroups and labs, and the special rapporteurs on extrajudicial executions and torture."[30]

One of the biggest limitations in the use of numbers based on uncertain data (Best, 2012; Landman & Godhes, 2013; Merry, 2016). Such numerical reasoning can be misleading since we do not know the reference frame or denominator that would guide meaningful comparisons. One interviewee used numbers in this kind of way as he described available data about the number of known nefarious groups: "… [What] I can tell you is that the difference between 250 and 7,000 is stark. The difference between 60,000 in the UK and 7,000 in the US is stark... That kind of

[30] Participant C4

supports your position right there."[31] However, comparisons such as these amount to what Best calls "apples and oranges." Overall, the analytical use of simple statistics, seen as having a secondary and complementary role, is also meant to advance an agenda. Advocates, however, tend to downplay the importance of quantification in their overall research and communication strategies.

4.1.3 Advocates also find numbers offer political "cover"

Only a few interviewees discussed their work in terms of following "rigorous" quantitative methods, whereby rigor here means abiding by well-established procedures of inquiry to arrive at scientifically valid inferences (King, Keohane, & Verba, 1994; Marquart, 2017; Satterthwaite & Simeone, 2016). Only one of the organizations discussed rigor in terms of holding themselves accountable to expert methodological peers as well as the human rights community. However, what's interesting is that even the more rigor-prone organizations emphasize the performative value of the numbers. G9 provides a specific example of a quantitative analysis being very useful to advocacy partners in abroad, "… not necessarily because the scientific findings are new or different, but because they add a layer of security. It's still really dangerous in most parts of [country X] to suggest going and investigating these cases, and so to be able to say, 'Hey, it's not us. It's the scientific model,' is actually really useful to them."[32]

While scholars make astute distinctions between the use of statistics for persuasive advocacy and statistics that are inferentially valid (Simone and Satterthwaite, 2016), we see how

[31] Participant U28
[32] Participant G9

inferential statistics also serve a rhetorical and persuasive function. Numbers clearly matter to advocates, sometimes for the empirical substance they convey, but most often for the attention and reliability numbers are seen to communicate. Overall, advocates describe the main role of numbers in advocacy to be performative. As G19 explains, “The numbers provide a very quick and easy way for the policy maker to both connect with the issue, and to give them evidence that making the advocacy change that we were requesting was the right thing to do, and giving them cover to do so.”

4.2 Advocates have a discerning sense of data limitations

Despite the value advocates ascribe to numbers for human rights advocacy, interviewees are not blind to the many issues that threaten the reliability of human rights data. All offer rich examples about the ways the data are difficult to access and collect, and how they are prone to under-registration and unevenness in their coverage and quality. In this section, I include examples of the many ways advocates talk about the limitations of human rights violations data.

Interviewees from international and U.S. focused groups talk about how social and political dimensions of violence create gaps or interruptions in data creation processes. For example, U28 talks about how the perpetrators deliberately hide their tracks, making data about their presence spotty. “They may go dark for months,” he says, which affects the passive daily and monthly monitoring counts. Another interviewee discussed how lack of trust in official institutions leads to data bias. “If you're from a community that's targeted by law enforcement, African Americans who are traditionally subjected to police brutality, you’re much less likely to trust law enforcement, and then less likely to report. Same with Muslim communities who are targeted in national security and terrorism investigations, much less likely to report to law enforcement...[furthermore] it's actually not mandatory for states to report this data."[33] Other groups describe how fear and coercion prevent reporting. For example, in the context of human trafficking data, another interviewee said: "Victims are having a hard time identifying as victims of trafficking because a lot of them are socialized to believe that terrible labor conditions are the norm or what should be expected."[34] Advocates working on sexual violence talk at length about victims resistance to disclosure given "stigma."[35] Many more examples of observational challenges where mentioned, including difficulties in learning about incidents from certain countries and challenges in verifying the veracity of online open source content. Overall, the anecdotes interviewees offers suggest that the data about "who did what to whom" is highly uneven in quality.[36]

From advocates in Colombia, obstacles to registration and instability in their data streams were also top-of-mind. For example, one participant talked about the obstacles to documenting paramilitary violence over time. He describes how it was only when the paramilitaries demobilized in the early 2000's and offered oral testimony did the Colombian state really begin to register past crimes committed by these groups. However, most streams of data rely on victim testimony. Having worked closely with victims for years, he describes the reticence he perceives people to have about reporting their experiences of victimization to the Colombian government agencies who they see as complicit in the violence. Even with the incentive for reparations offered, victims have said to him that "for political convictions, I am not about to take money in

[33] Participant G16
[34] Participant G15
[35] Participant G20
[36] "Who did what to whom?" is the name of a book by Patrick Ball (1996) where he lays out the units of analysis most human rights practitioners aim to count.

the name of my dead relative, and much less am I going to ask for it from the [state] perpetrator."[37] [38] He believes that even now, there is ongoing fear to report violations about the past: "I think what we know least about is the real degree of violence committed by [formerly-armed] groups post-demobilization… it's still too risky to denounce."[39] Overall, he worries that while available data about human rights abuses are inevitably incomplete, statistics based on available data are nonetheless prominent. He reflects on what he calls "a conceptual error that I think we've had here (in Colombia) as we manage statistics." He says, "We've always tried to bolster the statistic, ignoring or minimizing the under-registration or the fact that the figure I'm giving you is created in the context of war, so we confuse the data we know with reality." In this way, he recognizes the core tension between data production and the use of data on human rights violations.

Other Colombian interviewees talk about changes in their resources and their lack of access to certain regions of the country and how these issues affect the human rights data. One interviewee notes, "When I saw what we were able to monitor in 2009 compared with what we have the capacity to monitor today, to what we could confirm in 2009 to what we can confirm today, it's effectively 40% more."[40] According to another Colombian interviewee, "it is not possible to have a number on how many people have been assassinated …because everyone counts different things, and in reality, they are not comparable, furthermore one same source will

[37] Participant C8
[38] One of the largest databases of conflict-related victims in Colombia, the *Registro Unico de Victimas*, is collected with the purpose of administrating reparations benefits to eligible victims.
[39] Participant C8
[40] Participant C2

count one thing at first and then later count something else, and then another and another. They are not always clear about their criteria…they are not always transparent."[41]

All interviewees talk at length about data limitations; some go on to discuss the implications of this uneven data quality in the ways the data later get used. The main issue they bring up is commensurability. For example, G18 said, "You know, there are comparisons to be made but you can't just take our data and compare it to another data set and imply that they are directly comparable, and that's been done, I think, … with the particular intention of not only supporting the cause of that issue was, but somehow to diminish the [other] issues at the [same] time."[42] Another interviewee, after describing how her organization uses raw counts to compare whether the data increased or decreased from one year to the next, confessed, "I personally think it's problematic, it's problematic at least in a technical way because there are a lot of things that influence whether we can count more or less homicides, we've expanded [regionally], we have more people in the field, and because we're drawing a lot on information produced by others to know what's happening, and to count."[43] Still another researcher worries that when they focus too much on what is visible, they may forget to invest in issues or areas that are less visible: "…we need to make sure we're not just following the big, shiny ball in our research…."[44] Advocates own admission and discussion of data being unsuitable for making comparisons sits in stark contrast with the strong and confident expressions of the value of numbers. So, while the interviews confirm that practitioners strategically use the data to provide magnitude and patterns in data, they are at the same keenly aware of data limitations.

---

[41] Participant C1
[42] Participant G18
[43] Participant C1
[44] Participant G11

Overall, I find that the contrast between the value of number and the inherent limitations of data exists as a palpable tension for advocates. "All of that kind of confusion and nuance and imperfectness [in data], that's a great tension to point to," noted one participant.[45] Should data-driven arguments be made with such data? One interviewee struggled with this dilemma, "…this goes back to the tension I had mentioned between promoting the importance of data but then not wanting to tear it down so much, but then needing to tear it down."[46]

4.3 Advocates contend with tension between numbers and data uncertainty

One of my main findings from interviews is that advocates are aware of the data and uncertainty tension and that most attempt to address it to varying degrees. I find three ways that interviewees describe their efforts to mitigate the tension. The first is via technical efforts such as improving data collection and analysis. The second is to invest in educational efforts at their organizations, such as trying to improve staff's data literacy as they present and cite numbers. Third, and by far the most common considerations I heard, were pragmatic and somewhat instinctive assessments of the costs and benefits of using uncertain numbers without risking rebuke – a style akin to a "satisficing" approach. Satisficing is a concept introduced by Herbert Simon (1956) to describe the behavior or cognitive heuristics people use to arrive at satisfying and sufficient choices that meet acceptable thresholds when optimal solutions are not possible. As advocates described these pragmatic assessments, I heard some consistent themes emerge, so much so that I believe there may be some stable, if as-of-yet inexplicit, norms guiding advocates

[45] Participant U28
[46] Participant G16

to consider numbers "good enough" to use for the purposes of persuasive advocacy. In this section, I will briefly mention the technical and education efforts mentioned, and then focus on the ideas that I believe constitute a standard for "good enough" numbers for advocacy.

Briefly, with respect to improving data collection, all interviewees at organizations that produce data about human rights violations talk about trying to collect more and better data as a core part of their fact-finding work. They search for more data, especially from repressive contexts where international groups have a hard time visiting for field investigations, increasingly relies on open source content uploaded to social media by victims, perpetrators, and bystanders or shared through a variety of private platforms. Much effort is going into the challenges of verifying the veracity of online content.[47] In Colombia, organizations seek access to data about human rights violations via ad hoc agreements with outside organizations. The hope is that by collating many sources, they will have a more complete picture for their newest iteration of numbers reported. However, one interviewee pointed out that while collecting more data is valuable, it likely means her organization's monthly monitoring is not really commensurate over time, and each iteration may include a different set of underlying sources.[48] This is worth noting too because it highlights how if such changes to the data generation process are not explicitly communicated to end-users, this not only exacerbates the problem of data uncertainty, but gets lost as guiding interpretive frameworks for posterity.

A relatively small set of interviewees in this study talk less about collecting more data, and more about employing more "rigorous" statistical techniques to address inherent data

[47] Participant G11 & Participant G12
[48] Participant C3

limitations. By rigorous, I mean that the numbers and analyses presented to stakeholders are held to social scientific standards of quantitative research (Marquart, 2017; Satterthwaite and Simone, 2016). One organization, and somewhat of an outlier in this study, is dedicated to serving as a scientific partner to other human rights organizations around the world. Several others mention having formed partnerships with statisticians to produce some specific quantitative output for a specific campaign or issue. Representatives from two groups mention that their organizations' have taken steps in this direction, by hiring full-time in-house quantitative analysts. It is worth noting that in all of these efforts, the core tension with data uncertainty does not go away. What varies in the extent to which human rights advocates address the uncertainty in some explicit way, which I will discuss below. Because collecting more data or doing more rigorous analysis is resource-intensive and sometimes technically-prohibitive, a few organizations have decided that to take more of an educational approach to mitigating data uncertainty issues, akin to what Root (2016) describes as quantitative literacy trainings.

4.3.1 Advocates share consistent ideas about what constitutes "good enough" numbers

I listened to several interviewees talk through an implicit cost/benefit analysis and reason that the numbers that exist – whether they produce them or cite others' figures – are "good enough." Among the U.S. and international practitioners interviewed, doing more extensive quantitative work seemed beyond what they can aspire to with available resources. "It's just not feasible to do a broad-based quantitative study on the population that's been affected by a conflict because the scale that we're talking about is quite massive, and we're a team of about a dozen."[49] Some advocates explicitly distinguish that what they believe to be needed to achieve

[49] Participant G10

scientifically accurate numbers would require "way more energy than is certainly necessary for strictly advocacy purpose."[50] For advocacy, there appears to be a more relaxed data standard and conscious trade-off with producing scientifically reliable numbers. G17 put it this way:

> "[While] it would be really helpful to have those bigger numbers, I feel a little conflicted about it morally because to do a prevalence study, you're going out and finding and counting all of these people who are trafficked without actually tracking them and it takes a tremendous amount of resources to get that number. So, there's a small part of me that ... Well, there's a large part of me that doesn't want to spend that kind of influx of resources that would give us that number could do so much in terms of just adding more shelter beds in the US or more ... Even just training resources for prosecutors. There's so much you can do with that money."[51]

Among my study participants, I find practices akin to Jasanoff's proposal in action for the human rights advocacy context. As Science and Technology Studies scholar Sheila Jasanoff (2014) encourages for the role of science at the intersection of law and policy, legal scholars may be well-served to "set aside [the] some-what one-sided obsession with the quality of science," and instead ask "how science might best aid and advance the purposes of the law?" (p.1729). She distinguishes what may be "high-quality knowledge" from "serviceable truths," ultimately calling for some context-specific and pragmatic approximations to research products. In the human rights context, advocates recognize that there are scientific standards for the production of numbers, and while some aspire to those standards, most appear to accept a departure from those standards given the limitations of the data and the constraints of bringing attention to human rights abuses. The departure from ideal standards is not reckless; it aims to be principled.

Below, I analyze and find that advocates have pretty consistent and pragmatic ideas guiding their judgments about the numbers they say they are willing to use and how to present

[50] Participant G12
[51] Participant G17

them. Four ideas surfaced consistently: 1) numbers are considered temporary (not the final word), 2) conservative (low bounds), 3) empirical (based on data, not guesswork), and 4) communication of methods and data limitations should be transparent. This contrasts with prevailed dichotomous depictions. For example, Cohen and Green (2012) pose that human rights advocates' need to draw attention to human rights abuses creates an incentive for "drama," creating inherent incentives to producing "suspect facts" (p.446). They draw on the example of an unfounded statistic claiming that 75% of women had been raped during the Liberian civil war. They offer a theory whereby the incentive for "drama" in the short-term conflicts with the need for organizational credibility in the "long-term." As we will see, the advocates included in this study do not perceive their terrain under those stark parameters.

4.3.1.1 Four aspirations: empirical, temporary, conservative, and transparent

> "Very often we are able to give our best assessment, if you like, and it's not just guesswork. It's based upon the best information out there and available to us…the main thing is to not seek to exaggerate or inflate, to be clear about if there are methodology limitations or otherwise, but to present the information we have just as credibly as we can."[52]

In the quote above, G23 captures what I heard time and again: Advocates have low tolerance for guesses and falsehoods to advance their issues, they attempt to use available, empirical information and communicate it to audiences in what that will protect credibility. Just as Porter suggests that the power of numbers is less about accuracy and more about the appearance of rule-following, I find that advocates satisfice to this standard. They look to produce or cite data made on the basis of agreed upon categories and efforts to systematically

[52] Participant G23

classify events into them, even if that process ends up being messy and incomplete.[53] Many emphasize the importance for numbers to have a documentary back-up. In the words of one respondent, "we often have to say, at least X number, and we can provide names."[54] In Colombia, similarly, C8 noted that his organization's audiences want to know that any number provided is supported by names of the people counted. They said that it is for this reason that they opted out of seeking a more inferential approaches to estimating the number of people killed in the Colombian conflict. For advocates, it is safe to rely on data that is based on known incidents or people and that is produced by systematic classification efforts. While such data may have an uncertain relationship with the universe of violence in question, it arguably meets the threshold for what Porter's says is the essence of objectivity – i.e., distance from subjectivity. Thus, ongoing and systematic efforts to monitor and classify known incidents of violence and abuse, which may be far from meeting scientific standards of data reliable for inference, arguably are sufficient to create distance from subjectivity – i.e., to project objectivity – thus guarding against critiques of exaggeration and bias.

While interviewees stress that advocacy numbers must have some observational record behind them, they also stress the temporary quality of numbers. As one advocate put it, numbers are "by definition, a snap shot in time, … out of date almost immediately."[55] These temporary and incomplete counts are not intended as accurate portrayals, but rather to convey a sense of urgency to the situation. "[We] have no pretension of saying this is everything that happened in

---

[53] The production of categories is often highly contentious and political, the subject of lots of debate and disagreement behind the scenes. This is the subject of much debate among NGOs in Colombia. Merry (2016) describes the politics and ultimately the pragmatism of arriving at consensus categories in the context of human rights governance forums. However, once agreed upon and used for implementing counts, the politics largely disappear from public view. Recall that part of the uncertainty in data is how the ultimately agreed upon categories may include or excludes large swathes of the phenomena of interest.
[54] Participant G12
[55] Participant G22

the country, it only refers to what we know," explained a Colombian advocate.[56] The actual number is meant to trigger attention, as mentioned above. Importantly, it is seen as sufficient for this near-term and immediate context, and not as a consequential claim, either to scientific quantification standards or data-driven decision-making or long-lasting claims of magnitude. Still further, advocates overwhelmingly feel that as long as numbers are "conservative" and specifically described in their reports as minimums, they achieve their attention-impelling goal while staving off rebuttal and critique. "I think our position is very clearly that iterative improvements and measurements as we progress are better than not doing anything or trying to provide any type of answer."[57] Overall, this perception of numbers as temporary and conservative appears to ease the risk to organizational credibility or the possibility that numbers are found to be inaccurate at a later time.

Finally, as expressed in the quote by G23 above, a key idea expressed by most interviewees as an ideal for using 'good enough' numbers is credible communication. In this vein, most participants invoke the notion of transparency about methods and data limitations. "As long as the methodology is quite transparent, and as long as the source is quite straightforward in terms of the limitation of what they've done, I think journalists can then make a judgment call and decide whether a number is worth reporting or not."[58] However, while the notion of transparency came up often as key feature of 'good enough' numbers, further discussion revealed that transparency as an ideal is experienced more as a trade-offs than as a guiding principle, as I analyze below.

---

[56] Participant C1
[57] Participant G15
[58] Participant G23

4.3.2 While advocates express transparency as an ideal, it is rarely realized

As we saw in the Chapter 2 Literature and Theoretical Framework, scholars have been increasing their calls for more transparency within the human rights community about their production of knowledge – and specifically in relation to data-based claims about human rights violations (Cohen & Green, 2012; Greenhill, 2010; Price & Ball, 2015b; Root, 2016; Satterthwaite & Simeone, 2016). The calls for transparency simultaneously implicitly hold that conveying to audiences more information about how data are produced and analyzed – as well as disclose the limitations of the data – audiences will be in a better position to substantively comprehend the data and to scrutinize its reliability. These same scholars also sometimes make sweeping statements suggesting that if advocates offer more transparency to audiences, they will benefit with increased credibility for their organization.

Aligned with these general ideas, interviewees in this study also express transparency as a high-level aspiration for themselves and as important to assess fact-based human rights claims produced by others. When asked to explain their approach to such transparency, most spoke about their increased investment in methodological description. However, when probed specifically about whether they value and practice disclosing specific information about the limitations inherent in data on which quantitative claims rest, we begin to get a better sense of how the ideal of transparency about data production and uncertainty in human rights statistics becomes complicated by various incentives and disincentives participants perceive to be inherent to the implications of transparency for advocacy work, as analyzed next.

4.3.2.1 Methods sections are often high-level

Most, but not all, participants report that their organization has invested effort to be more transparent about their research process. Primarily, they talk about enhancing the methodology sections in their published reports. For example, one advocate I interviewed said that the organization's explanation of their methods "is much more robust… than we used to have, and saying what we do and what we don't know about the people that we've interviewed."[59] Participants describe pretty consistent and formulaic approaches: "Look, here's where I was able to travel. Here's what I was able to do. I spoke to X number of people or whatever."[60] These kinds of sections are primarily aimed at communicating to audiences that there was a research process to support the reported information.

Taking a closer look, however, many reveal that the reason methodological transparency is important is often more symbolic than substantive. Participants talk about methods sections as being important signals of credibility. G23 captures it in one phrase: "[Methods] are important in terms of establishing the credibility, I guess, of our work, but it's not necessary to the key advocacy message."[61] Providing substantive methodological detail about the data production was described as relatively marginal and sometimes inconvenient. One Colombian interviewee talked about the choice to exclude his survey methods from a public facing report because lay audiences would not be interested, "the report tried not to be too technical about all the sampling. I added [details about methodology] to the presentation as a technical appendix, but it doesn't

[59] Participant G22
[60] Participant G18
[61] Participant G23

appear in the publication."[62] A U.S. advocate echoed this sentiment. "Well, you know, we aren't releasing pure view scientific reports. That's the first thing. So it shows up less in the sense that we don't have a methods section in which we address the weaknesses…. Doesn't mean that all of our methodology stuff doesn't really hold up, it just means at times we're gonna have different tactics towards a similar goal."[63] Much of this sentiment appears to be rooted in an assessment that audiences just want you to get to the point. Of note, Lupia (2013) analyzes the communication of science in politicized environments and finds that science communicators often overestimate their audiences capacity to pay attention to scientific findings and methodological underpinnings. In the human rights contexts, it appears that advocates anticipate audiences' capacity and (little) interest in this information more than some academics.

With regard to the substance of methodological description, some participants actually feel stifled by what they perceive to be a lack of interest in methodology by their audiences. "I think the demand for knowing about the methodology is less than how much we would actually like to talk about it."[64] A U.S.-based advocate said, "I don't know that there is a heck of a lot of interest, certainly not by the public. I find, frankly, among policymakers, that the interest around methodology frequently comes from people that are already deeply skeptical of the analysis that you're doing…"[65] However, with respect to this skepticism, there is a sense that political detractors will take a more political tack than a methodological one. As one respondent said, "Yeah, I think some of the more unscrupulous governments and state actors don't really care, to be honest. They don't like what we're saying and they want us gone basically. That's why we'll

[62] Participant C6
[63] Participant U28
[64] Participant G18
[65] Participant U27

often get accused of fake news and things like that... Even heads of state have said that about some of our reporting, and they just don't even want to enter a dialogue about our methodology, and why we think they should do more to account for X, Y, and Z."[66]

As a whole, I find that interviewees discuss their consideration of providing methodological description in relation to short-term pragmatic trade-offs. From interviewees' perspectives, it appears that vis-a-vis committed adversaries, more transparent methods will rarely serve as a bulwark against critique, and with the media or more sympathetic audiences, there is little interest in substantive methodological discussion. The relatively small community calling for more transparency is making appeals to higher research standards (which advocates do not appear to think are required strictly for advocacy work) or to consequences on credibility, which as we will see, does not appear to be at stake if advocates follow "good enough" standards in providing quantitative information for advocacy or when the numbers play a relatively small complementary role. In sum, while methodological transparency is seen as an ideal, it is not considered to be required for advocacy, and can at times be at odds with it. Thus, there appears to be little incentive for advocates to develop this part of their reporting beyond current practices.

Unsurprisingly, when interviewees put themselves on the receiving end of human rights reports published by others, some find methods sections to be generally short and lacking detail. One respondent complained "…in their reports, they normally include methodology and it's very frustrating when I'm looking at it and trying to understand what's going on and I don't quite understand it."[67] Another respondent similarly described that even when working in partnership

[66] Participant G10
[67] Participant G13

with local human rights groups, “it’s really difficult to get the groups to articulate in a way that's useful to us how they're collecting the data.” These short-term trade-offs in providing methodological detail, while functional and pragmatic in immediate advocacy contexts, could have long-term implications for how the data is used and interpreted in future use contexts, such as comparisons of data over-time, decision-making, and historical truth-seeking efforts. I will return to some of the potential downsides of such limited methodological transparency below.

4.3.2.2 Advocates describe deterrents and risks in presenting data limitations

Advocates distinguish the description of methodological procedures from more specific detail about the relative strengths and limitations of the data collected and used in other parts of reports or messages. In fact, interviews reveal that communicating specifically about data uncertainty – the limitations of data used in relation to quantitative claims they are used to make – is far more complicated.

Several interviewees talk about the intention to provide transparent communication about data limitations. Excerpts from my conversation with one interviewee on the topic echo the sentiment of some of the groups that take pride in their research efforts and ethics.

> “Well certainly I think every research team wants to convey [data limitations]... …I think the problem becomes that in translation to what the public is interested in, those caveats lose salience or lose priority because that's not typically what the focus of a lot of these interviews become…. in communication to the broader public, I think people start to care less and less about those nuances…the best you can do is push for caveats, indicating this is under-representation, children are not well-represented, that [a certain region] is not well-represented due to data availability issues, but honestly, it's really not easy to get the general public to even care about these caveats.”[68]

[68] Participant G15

G15 went on to talk about how she encourages her audiences be more critical information consumers: “Don't be so eager to have this one-line soundbite with just the number without any context.”[69] Another interviewee was similarly disheartened in what she perceives as a lack of appreciation for expressions of uncertainty. Referring to her presentation of statistical estimates to a group of policymakers, G9 recalls, “They wanted no discussion of ranges, no bounds, no uncertainty, no ... They were just like, "It's a number. It's the number of names that have been written down. Just tell me how many it is." And I was like, ‘Here's the range.’”[70]

A much larger set of interviewees focused less on the lack of demand for data uncertainty information, and more on their own dilemmas with providing such information. Participants raised three main concerns with communicating data uncertainty to audiences: such information can be 1) unhelpful and uninformative, 2) unreliably read, and 3) cast the source as an unreliable information producer. All three of these issues, raised more as intuitions by respondents than as demonstrated outcomes, nonetheless resonate with scholarly findings about the potential perils of communicating uncertainty.

The first common theme was the idea that being transparent about data limitations can be unhelpful and uninformative, creating more noise than signal for their audiences. This echoes general concerns that people are poor at processing uncertainty information (Kahneman, 2011) and that policymakers want compressed sharp information (Lupia, 2013). For example, C4 overheard a conversation among some ambassadors complaining about missing data in his

[69] Participant G15
[70] Participant G9

organizations' violations lists, calling the data "'useless…filled with NNs and cases with no location information. We don't know if a record is one, two or three people.'"[71] What to do with missing data, and whether to convey it, is a common issue. G18 worries about how to use and convey such missing data when making quantitative claims. "Often we just leave out unknowns in graphic representations because sometimes there are so many that it would distort the data. Then sometimes we leave them in, so you even see that in one graphic representation where we've chosen to leave out and in another to keep in…that's where we put the asterisks and say 'Of the number known.'" In this case, we see G18 grappling with the idea that the data visualization may be misleading in the first place, and that the caveat tends to be uninformative. Overall, there is a sense that caveats such as these can equally create confusion because they do not add information about its degree or implications of the missing data. On the topic of under-registration in datasets of victims, another participant reasoned it was best not to raise the issue precisely because they do not have a useful way to help audiences make sense of it, "No, we only speak of what we know, we do not touch the topic of under-registration because we have no idea how to measure it, we would only be speculating, so if we could measure it, we would, but without that, we do not even mention it."[72] Overall, there is a sense that pointing to data uncertainty without offering some additional insight can be less, rather than more informative.

Another potential risk participants see is that information about uncertainty will be unreliably read. One participant said she received positive feedback about her organization's decision to include caveats, with comments such as "it's very interesting that you make that

[71] Participant C4. NN is the common way to refer to an unidentified body in Colombia, a corpse with "no name."
[72] Participant C2

clarification because now one knows better what to stick to in principle."[73] More often, however, interviewees worry that uncertainty will be negatively read as them downplaying the relevance of the cases that are known, as another interviewee pointed out: "so when one puts too much emphasis on the under-registration it's like leaving a subtext that one thinks the reported number of victims are just a few, no?" Perhaps another interviewee best captures the fine line advocates are considering about the possible ways in which uncertainty information will be read "…it can be very positive or very counterproductive, we have to look at the context. If demobilized guerillas are being killed and we come out and say well the stats are very uncertain, one almost looks complicit…so we have to look very carefully at the context.. one thing is whether there is uncertainty, and a very different thing is that everything is OK… there are a whole range of distortions that are based on statistical uncertainty to say 'nothing is happening here.'"[74] These intuitions about information being negatively read are consistent with empirical evidence in other contexts, whereby verbal expression of uncertain care unreliably read (Wallsten and Budescu, 1995), or are understood with inverted valence or 'boomerang effects,' effects (e.g., regarding climate policies and with the presentation of weak forensic evidence in courtrooms (Martire et al., 2014).

Finally, a last common concern is that uncertainty will make the data appear unreliable, which in turn makes the data producer appear less valuable. As C8 put it: "to sustain an NGO, it is not too profitable for someone to say, 'I present these data, but let me clarify, the under-registration can be terrible, we're in the context of war, no?! The presenter must say, this is reliable, it's valuable information. One doesn't present an product to also give all the 'buts,'

[73] Participant C3
[74] Participant C7

maybe it's an issue of survival of the data producer, that's what's at play, no?"[75] And, at the most extreme, one interviewee shared his concern that uncertainty information could be used as political ammunition, "it puts you at a rhetorical disadvantage to go and say well these are all the ... 'cause you're basically putting a neon sign towards your detractors to say, this is why this is bullshit…"[76]

These various issues, from the lack of demand for uncertainty information to the disincentives and potential risks for supplying uncertainty information, are not unique to the human rights community. They echo disincentives expressed for communicating uncertainty in many politized contexts and even disincentives to publish negative results in academia (Nosek, Spies, & Motyl, 2012). The point here is that they are present issues that advocates grapple with as they consider the ideal of transparency for their immediate advocacy contexts. Being transparent inevitably creates a widely-shared dilemma, as reflected in the words of G16, "I would not be in favor of leaving that information, which I think is very important, out of advocacy work for the sake of making it more compelling, although, not to say that I would disagree if someone were to come out otherwise and try to downplay that caveat in service of some advocacy goal." G16 captures the tension between the desire to be transparent about uncertainty and the potential, multifaceted deterrents for actually being so. The foreseen cognitive and political resistance leaves an unclear path for advocates as to how to handle the question of transparency about data uncertainty. What comes across strongly is that the consideration of the trade-offs mostly sound like short-term pragmatism rather than long-term considerations about how this information will be used in contexts beyond immediate advocacy.

[75] Participant C8
[76] Participant U28

Next, I look at a small sample of how methodology and data limitations are actually conveyed in a small sample of Human Rights Watch (HRW) publications.

4.3.2.3 Preliminary review of HRW reports

To deepen our comprehension of how advocates communicate numbers, methods and data limitations in their publications, I conducted a pilot study of a small random sample of twelve HRW reports, six from 2012 and six from 2018.[77] Given the reduced time I had to conduct this complementary, yet initial analysis, I focused on reports written by HRW because they are arguably one of the most influential organization producing regular human rights reports.

In the analysis of this small sample of reports, I find HRW's communication practices to be quite similar to what advocates are sharing in interviews and to what has been found in previous similar studies (e.g., Satterthwaite and Simone 2016). First, while HRW is primarily a qualitative research enterprise, almost all reports include quantitative claims (10 out of 12). These quantitative claims function just as interviewees suggest – either as an empirical way to demonstrate that incidents are occurring at a high frequency or to complement a more qualitative argument in the report, drawing on data that show a trend exists.[78] Methodology sections tend to be generic, briefly outlining "what was done" to collect the information used in the report. Limitations and caveats mentioned in this section are in relation to overall research limitations

[77] Many thanks to my research assistant, Alexa Patrick-Rodriguez for her help coding the reports. The methodology for this study in included in the methods section above. As a reminder, it's a small study given the limited time to conduct it.
[78] An example of an "a lot" statement looks like this: "Between 1996 and 2011, city police made 586,320 arrests for possession of marijuana in public view in violation of New York Penal Law § 221." (HRW 2012, p.10). An example of a "bolster claims" statement looks like this: "An analysis of marijuana arrests in the period 2004-2008 revealed 48 blacks arrested for marijuana possession for every 1,000 in the population, 24 Hispanics arrested per 1,000, and 6 whites arrested per 1,000" (HRW 2012, p.13).

(which tend to be primarily qualitative in nature), rather than addressing specific flaws in quantitative data used in other parts of the report. Overall, these findings support the expressed sentiment that methods sections are more general and symbolic than substantive.

In other sections of the reports where specific numeric claims appear, rarely does one find language about the limitations to the quantitative claims (only 1 out of 12 included something resembling a caveat statement). The common way reports note the lack of precision to their numeric claims is to include 'approximate' statements – phrases such as "more than" or "at least" (seen in 8 out of 10 reports). Overall, numbers in reports are communicated as confident claims, revealing little to nothing about data uncertainty. While these are preliminary findings based on a very small sample of reports and should be strengthened and validated with a larger, more representative sample, they are consistent with interviewees' sentiment whereby "good enough" numbers with low transparency about the data back story is stable practice in human rights advocacy reporting.

4.3.3 Methodological scrutiny is selective when advocates cite "authoritative numbers"

For those organizations that rely on citing numbers produced by others, interviewees have a pretty consistent sense of where and how they look for numbers that they are willing to cite. For example, G26 said "…we do use numbers that have been collected by others, authoritative numbers, that's part of our advocacy all the time. I mean, I think it's hugely important."[79]

[79] Participant G26

A few respondents return to notions of transparency, or methodological disclosure, as being important. As G14 put it, "the types of sources that we would be looking at and be willing to cite as authoritative in our work would typically have a methodology because they would have pretty high bars in terms of the quality of their work." But through my multiple conversations with advocates, it became clear that for the most part, methodological scrutiny is selective; just as most people, advocates use heuristics to determine source credibility (Chanthika, 2006). Methodological detail becomes important when the available numbers are produced by organizations that are relatively unknown to them. "I think the methodology matters more when it's your opponent's data than when it's your ally's data. It probably shouldn't be the case, but I think there's a bit of an echo chamber and sort of pumping out and trusting that those of us in this are all legit and have the same amount of rigor in our work. It's just not the truth, despite the perpetuating ethos or just environment."[80] In Colombia, C8 expressed a similar conviction, "One things that happens when we build reputations … is that society does not ask itself much about the route…Sometimes the findings is accepted or not depending on the reputation of the data producer." So, this logic follows Porter's theoretical explanation about the authority of numbers. He posits that when personal or institutional trust is low or unknown – in this case political trust – then trust in numbers has an opportunity to exert its power. In the case of human rights numbers, advocates' enhanced scrutiny that the rules of objectivity have been properly implemented (in this case, a review of methodology) when other means of trust are "distant" – in this case, distant socially and politically – is aligned with Porter's theoretical explanation. It also reveals how at times, transparency can be drawn upon for accountability, but may only be done so when other factors for sources' credibility are weak.

[80] Participant G21

By far, interviewees say they prefer to cite United Nations agencies or affiliates, which they say they trust, and overall feel will be well-received and pose little risk of rebuke. “Yeah, for the UN, … we don't necessarily look in depth at their methodology every time they put out a new report, we sort of assume that they're using the same methodology and that it's built on what they've done before.”[81] However, many concede they know little about the methods of such organizations. At least four interviewees independently said they have found some UN numbers to be “flaky” and that inclusion of methods for the numbers they put out are “hit or miss.”[82] One interviewee confessed, “In some cases people use very authoritative numbers that are actually very weak and very unconvincing.”[83] It appears that if advocates can offset accountability for uncertain numbers to a ‘reputable’ organization, their own credibility can be protected. Elsewhere scholars have found that source credibility is primarily determined by *perceived* common interests and *perceived* relative knowledge, which offers analytic traction here. According to Alston and Knuckey (2016), dominant forms of trust in the human rights community usually come from shared commitments, values, and norms and less from social scientific disciplinary practices.

Citing numbers that they accept as being weak is seen in a related way among Colombian and U.S. respondents. Several said they prefer to cite numbers from “official” institutions – usually government institutions – even if they know well that their numbers are incomplete and potentially biased. “You provide a figure that the state is telling you about its own behavior, who is going to refute that? To argue publicly, it is a powerful tool.”[84] They feel they can avoid

[81] Participant G22
[82] Participant G14
[83] Participant G26
[84] Participant C3

rebuke and advance their agenda more if they can work with government numbers. Overall, just as advocates satisfice in their assessment of what numbers serve as “good enough” and to what (limited) degree they need to convey data uncertainty to audiences, they use pragmatic heuristics to assess which external numbers they can safely cite without having to invest too much time vetting methodological rigor. Much of what they share about these practices to be intuition and tacit knowledge about costs and benefits, rather than conclusion to which they have arrived as through much deliberation.

Advocates appear quite aware of several of these tensions operating at the intersection of the production and consumption of quantitative measures of human rights violations, and their approach to dealing with the tension is inherently pragmatic. Among their biggest concerns, as one advocate put it, “people are pretty concerned about putting in something that can be rebutted.”[85] However, based on these interviews, I find that irrespective of some degree of transparency about data uncertainty (and even sometimes thanks to a lack of transparency), using “good enough” numbers is perceived as effective and posing little risk to advocates in the way there are currently using them. Granted, these norms are quite permissive; they allow “flaky” numbers to pass, including potentially misleading comparisons by advocates themselves. They may also allow false, exaggerated or doctored numbers to pass. While they might be more careful if they anticipate challenges to numbers, interviewees were hard pressed to find cases involving themselves or others where political detractors focus their critique on the uncertainty of published numbers, their methods, or their limitations.

[85] Participant G19

As we saw in case of the CNMH's *Enough Already!* report, political considerations tend to outweigh methodological ones – allies rarely call each other out. While I did hear some vague mention that people see suspicious numbers occasionally, they share that those tend to get addressed among peers quietly. For example, "So for us to verify the numbers, it is a kind of cooperative exercise, where if we don't understand, if something doesn't add up, we try to go to the person who prepared that report for the government first and say, "Hey. Do you realize this calculation is off?" If we can't reach them, we try and meet the diplomats, brief them, so that they can send a message back to capital and say, "What's this?" We try and compare it to what we know from past reporting and from what other countries have produced. But there's not a lot of transparency…"[86]

4.4 Conclusion

In conclusion, we see that for a few scholars and organizations, numbers are a way to "speak truth to power," thus they aim for numbers to be rigorous, scientifically defensible, and transparent. This is especially so for high-stakes forums such as truth commission, courtrooms, and data-driven decision-making. G9 discusses when she considers it worthwhile to invest in producing rigorous numbers:

> "That's the million dollar strategic question that we're struggling with right now because internally, we're not convinced that that's ever particularly useful. But we recognize that [good enough numbers] are often what you need to open the door to get somebody to write a headline or a press release or have a meeting, to then get to what we think are the significantly more meaningful comparisons about all the usual stuff that we would do. What's the change in patterns of violence over time or space or ethnic group? Most people we find, can't figure out how to start there."

[86] Participant G24

However, for the majority of advocates in this study, numbers serve as a means to a more short-term, instrumental and low stakes end – usually to bring much needed attention to important issues. The fact that advocates disseminate uncertain numbers relatively freely does not necessarily mean those numbers are deceptive, as those who paint stark dichotomies of "good" and "bad" numbers might contend (Best, 2012). In the immediate term, the onus and liability to get accurate numbers, or even to present them more carefully is largely not a main worry for their use in advocacy contexts. As we will see in Chapter 5 Experimental Survey Findings, when numbers are used in a consequential scenarios like decisions-making, those numbers can shape outcomes significantly, and only highly directive messages of data uncertainty can peel away at their anchoring power.

Chapter 5 Experimental Survey Findings

5.1 Introduction

The main concern in this chapter is that over-reliance on uncertain data by decisionmakers. Even if it is the best available data, data with significant uncertainty can nonetheless provide a false sense of certainty to decision-makers. As we have seen in Chapter 4, groups trying to call attention to injustices and inequality put numbers based on uncertain data out into the work to trigger attention, and while advocates in my interview study reason that such numbers may come with relatively little risk in the advocacy context, such numbers about social issues nonetheless travel to other contexts (Andreas and Greenhill, 2010). Some scholars have expressed specific concern about partial and biased data and numbers having potentially harmful impact when taken out of the original context of production and use, and then used to influence real-world decision-making. In some examples, biased data has been shown to amplify racial and socio-economic bias in criminal justice contexts, enable funneling of attention and resources to only the most visible problems, undermine the credibility of human rights institutions, and privilege certain narratives over others (Cohen & Green, 2012; Kruger et al., 2013; Lum, 2017; Reydams, 2016; Starr, 2014).

As we have seen, scholars from across disciplines are calling for more transparent communication of data limitations to audiences, especially when important decisions are at stake (boyd & Crawford, 2012; Greenhill, 2010; O'Neil, 2016; Root, 2016; Satterthwaite & Simeone, 2016; Tufekci, 2014). Given the expertise, cost and difficulties often involved in calculating statistical bounds of uncertainty (Fariss, 2014; Kruger & Lum, 2015; Price & Ball, 2015b; T. B. Seybolt et al., 2013), an alternative recommendation to practitioners is to "confront limitations and bias [in data] through language" (Root, 2016, p. 364). While calls for transparency are often rooted in ideas of accountability (Ananny & Crawford, 2018), calls for such transparency in the human rights community are also rooted in the idea that by expressing uncertainty, the communicator offers complementary information to audiences to buffer them from interpreting numbers with false precision or inaccuracy.

But as described in detail in Chapter 2 Literature and Theoretical Framework, cognitive science and science communication scholarship reveals that communicating uncertainty effectively is not straightforward. Research from disparate scientific domains provide scattered insights on the challenges of communicating uncertainty. First, communicating uncertainty is often considered a risk in politized environments, and we have seen that human rights advocates in this study anticipate and attempt to minimize such risks. With respect to tobacco, climate policy and other public policy issues, Oreskes & Conway (2010) show how politicians distorted scientists' expression of uncertainty to sow doubt in favor of their agendas. Other scholars advocate that it is important to help policymakers to be better consumers of uncertainty and to push back against "incredible certitude" in policy decision-making (Fischhoff, 2012; Manski, 2018). In cognitive experiments testing specific implementations of uncertainty expressions, studies show that better decision outcomes can emerge from numerical expressions of uncertainty than linguistic ones (Budescu & Wallsten,1987; Martire et al., 2013; Joslyn & LeClerc ,2013; Budescu et al., 2012); that making explicit recommendations to decision-makers in the face of uncertainty is likely to lead to improved outcomes (Joslyn & LeClerc, 2013; Joslyn et al., 2009); that "hedges" can increase perceptions of credibility of information providers

(Jensen, 2008) or at least not lower them (Retzbach & Maier, 2015), but can also weaken the perceived value of evidence in court (Martire et al., 2014; Kadane & Koehler, 2017). To support domains where uncertainty is seen to lead to better decisions, some scholars are exploring the best ways to visually convey uncertainty (Fernandes, Walls, Munson, Hullman, & Kay, 2018; Greis, Hullman, Correll, Kay, & Shaer, 2017; Hullman, 2016).

Yet even in these other domains, the findings on communicating uncertainty are isolated and sometimes contradictory. In human rights contexts, while there are increased calls for communicating uncertainty in data to audiences, how to communicate uncertainty effectively remains an open challenge and the impact on decisions and perceptions of source credibility remain unknown. As seen, I began addressing this issue by interviewing human rights advocates on their views about using uncertain numbers and expressing data imitations to audiences. To gauge the impact of data produced in one context and then moved to decision-making context, I designed three human rights vignettes, accompanied by one of four different data uncertainty messages. I drew on science communication theory and human rights communication practices to design each message type (described below).

As described in Chapter 3: Research Methods in detail, I implemented a between-subjects randomized control trial with 970 college graduates that included 6 experimental conditions (2 controls and 4 treatments) which I tested across three different vignettes (6 x 3 design). More specifically, the purpose of the experiment was to collect evidence of whether there was any basis to thinking minor changes in uncertainty messages would shift outcomes in meaningful ways. Findings (presented in this chapter) confirm that when solitary numbers are presented with no uncertainty message, decisions significantly shift in accordance to the numeric information (compared to the same scenario with no numbers presented). Unsurprisingly, this confirms the strong anchoring effect of the numbers, what Kahneman (2011) says is "one the most robust and reliable results of experimental psychology" (p. 119). The most interesting finding of this study is that only one message type consistently leads decisionmakers to attenuate their decisions in the face of numeric information – an uncertainty message I call "expert interception." In these messages, designed on the basis of insights Joslyn & LeClerc (2013), the presenter of information anticipates the anchoring effect of the number and the tendency of audiences to interpret numbers as representative of real-world trends (i.e., to not foresee or consider the impact of partiality and bias in available data) and explicitly indicates the dangers associated with doing so. The study also finds evidence in support of the idea of what Ball (2016) and others have called a "caveat fallacy," which essentially refutes the idea that any linguistic hedges attenuate the impact of the anchoring effect. Finally, while science communication scholars have found that signaling uncertainty in scientific findings may either lead people to consider the information provider more trustworthy or at least not harm perceptions of trustworthiness (Jensen 2008; Retzbach and Maier, 2015), this study found trust scores vary with the way different uncertainty messages, at times remaining unaffected, but with the most "effective" message type (expert interception), trustworthiness slightly decreased. After presenting these findings in detail, I discuss some of the implications of these findings (along with some of their limitations) for scholarship and human rights practice.

5.1.1 Sample Description

Respondents were 57% women and 43% men. Respondents age was, on average, 39 years old, with the minimum and maximum ages at 21 and 82 years old, and the standard deviation at 11. Respondents took an average of sixteen minutes to complete the full survey. Data were collected via a customized data collection instrument on Qualtrics. I pre-tested the data collection instruments in March and August 2018 to refine vignettes and treatment language, mostly for reading comprehension. To determine the average effects presented below, I conducted OLS regression analyses.

5.2 Main Findings

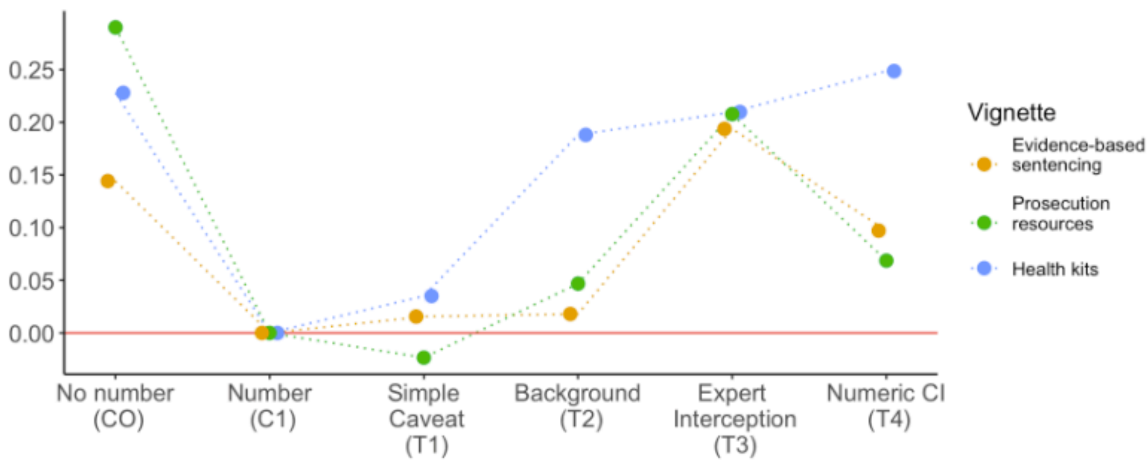
What is the impact of different expression of data uncertainty on decisions and perceptions of source trust?

5.2.1 Decisions

To describe findings, I make the C1 control condition the main point of reference – that is, where numbers are presented to the decision-maker, but no uncertainty message is included. Figure 2 presents 6 (experimental conditions) x 3 (vignettes) results for decision means. To make results comparable across vignettes, decision means are standardized with respect to C1 as follows: $\boxed{(\bar{X}_{C_1} - \bar{X}_{t_i})/\bar{X}_{C_1}}$. The horizontal axis presents the 6 experimental arms, and the vertical axis presents the percent change in the mean with respect to C1. The standard errors are included in vignette specific results in Figure 3 and in the regression tables, below. The challenges of plotting uncertainty in data visualizations is a whole other topic of research (Hullman, Qiao, Correll, Kale, & Kay, 2019).

The three strongest findings that relate to the hypotheses are as follows: 1) the impact of including a “number, no uncertainty” (C1) versus “no number” (C0); 2) the consistently significant impact of "expert interception" (T3) aimed to intercept the tendency to view numbers as representative of trends in reality, and 3) the consistently null effect of the "simple caveat" (T1) to sway decisions with respect to "number, no uncertainty" (C1). I discuss each of these in detail below.

Figure 1: Decision Mean Standardized relative to C1



** Light dotted lines just guide the eyes by vignette, to see differences across the discrete treatment conditions. For example, for the "evidence-based sentencing" vignette (orange), the mean for C0 has a 15% difference from the mean of C1.*

5.2.1.1 The Power of Numbers

Table 4: C0 (No Number) vs. C1 (Number) Decision Means[87]

	Evidence-based sentencing		Health kits		Prosecution Resources	
Treatments	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Number vs. No number	0.70 (0.37)	0.058	-62.05 (9.72)	**<0.001**	-7.89 (1.56)	**<0.001**
Observations	378		378		341	

I find evidence in support of hypothesis 1 – that the inclusion of "numbers, no uncertainty" (C1) significantly shifts decision means with respect to the baseline where "No Number" (C0) were provided to guide the decision-maker.

As seen in Figure 3, Numbers (with no uncertainty - C1) consistently serve as a signal for decision-makers. In the "evidence-based sentencing" scenario, we see that including a risk score (C1) leads, on average, to an 8-month longer sentence than not including the score at all. This finding is significant at the .1 level, just shy of being significant at the stronger .05 level.[88] In the allocation of "health kits" to rape survivors, including Numbers (with no uncertainty - C1) of reported victims in two regions leads towards a proportional allocation of kits. In the "prosecution resources" case, as expected, the numbers of documented victims by ethnic group also led people to adjust towards a proportional allocation. Interestingly, in this case, people did not go as far as the numbers would indicate: a purely proportional allocation would have been

[87] N = 378 are respondents assigned to C0 and C1 experimental conditions. For the prosecution resources, I had to exclude 37 records due to data entry errors.
[88] I am exploring why the shift between C0 and C1 is less pronounced than in the other two vignettes. Early exploration leads me to believe there is a tempering effect due to correlations between the sentencing decisions and people's political views of the criminal justice system.

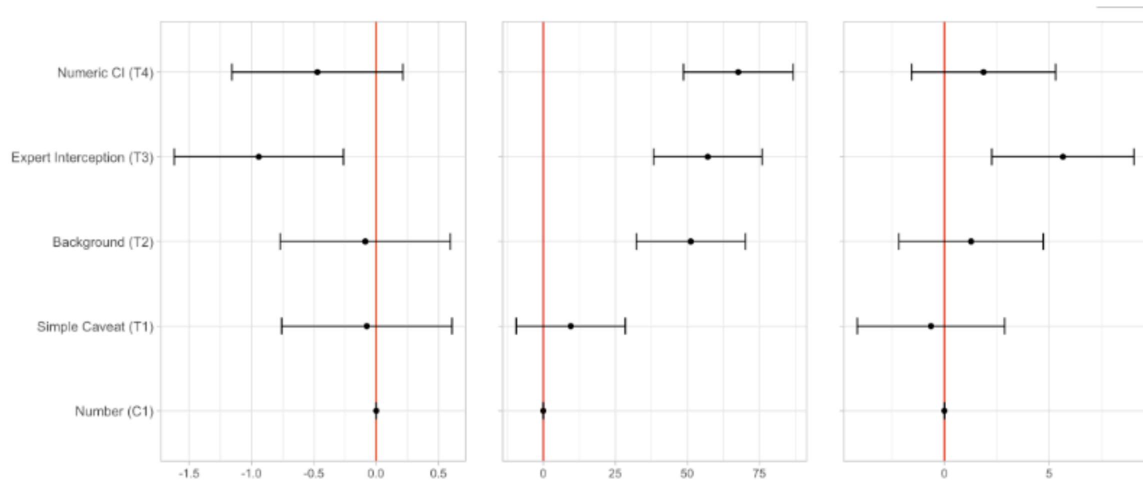
given 90% of available resources to one group and 10% to the other yet respondents only adjusted to a 65/35% division.

5.2.1.2 Effective and Ineffective Uncertainty Messages

In a Bayesian framework, we might consider the mean of C1 as people's priors about how to evaluate numbers – what people's allocation would be, on average, in the presence of numbers, no uncertainty. In this case, treatments effects could be understood as how people update in the face of uncertainty. The communication goal is to attenuate the priors (C1). In the specific vignettes, information providers want to attenuate the effect of the numbers because they are concerned that risk scores can amplify bias in crime data ("evidence-based sentencing" case), that resources may disproportionately favor the most visible victims ("health kits" case), and that uneven reporting of victims across ethnic groups may led to impunity for the perpetrators of victims with less documentation ("prosecution resources" case). So, how do people update in the face of these uncertainty messages?

Figure 2: Decision Means per Vignette

“Evidence-based sentencing” (left) *“Health Kits” (center)* *“Prosecution Resources” (right)*



Vignette-specific results with the point estimate and 95% confidence intervals of the effect size for the contrast between C1 and each treatment.

Table 5: Decision as Dependent Variable

C1 vs. all treatments:

	EBS	EBS ctrls	SVN	SVN ctrls	PRA	PRA ctrls
Predictors	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*
Number (C1)	4.86 *** (0.25)	4.07 *** (0.30)	272.27 *** (6.84)	284.68 *** (14.61)	27.21 *** (1.28)	29.94 *** (1.40)
Caveat (T1)	-0.08 (0.35)	0.04 (0.33)	9.54 (9.66)	9.51 (9.60)	-0.64 (1.79)	-0.51 (1.76)
Background (T2)	-0.09 (0.35)	-0.02 (0.33)	51.20 *** (9.65)	52.92 *** (9.60)	1.27 (1.76)	1.67 (1.73)
Expert Interception (T3)	-0.94 ** (0.35)	-0.86 ** (0.33)	57.10 *** (9.60)	57.96 *** (9.55)	5.65 ** (1.73)	5.81 *** (1.70)
Numeric CI (T4)	-0.47 (0.35)	-0.40 (0.33)	67.68 *** (9.70)	67.42 *** (9.64)	1.87 (1.75)	1.89 (1.73)
Expert (PRA)		0.57 *** (0.13)				
Political View		-0.74 *** (0.09)				-1.28 ** (0.46)
Age				-0.26 (0.27)		
Expert (SVN)				5.18 (3.26)		
CRT				-11.06 ** (3.60)		-2.73 *** (0.63)
Observations	952	952	952	952	867	867
R^2 / adjusted R^2	0.011 / 0.007	0.097 / 0.091	0.077 / 0.073	0.092 / 0.085	0.019 / 0.015	0.053 / 0.047

• *p<0.05 ** p<0.01 *** p<0.001*

In Figure 2: Decision Means per Vignette, we see the effects of each treatment in contrast with C1 (Number, no uncertainty). The red line indicates the mean at C1, and the intervals are the standard errors of the effect size of each treatment with respect to C1. Significant effects at the .05 level are those where standard errors do not cross the red line.

Of the four treatment conditions, T3 (Expert interception) was the only type of uncertainty expression that consistently changes decisions in the direction indicated by the uncertainty message (drawing on the context-specific expertise of the communicator), and away from the allocation suggested by the numbers provided. Recall that T3 is designed on the basis of insights from Joselyn and Leclerc (2013) on communicating uncertainty in weather forecasts. These messages have two key features: they 1) foresee and intercept common cognitive misinterpretations (thus allow people to think more carefully), and 2) relate the message to the decision task. In these scenarios, I designed language to intercept the common misinterpretation that data represents a reliable pattern in reality. It is encouraging to see that the effects of this message type hold in this different domain and across vignettes. This can be seen as strong evidence in support of hypothesis 3.

Recall that T1 (Simple Caveat) was intended to be a signal of caution rather than an informative message (Crismore & Kopple, 1988). As predicted, T1 had no impact on decision-makers in any of the vignettes, providing strong evidence in line with hypothesis 2. This supports the idea of a "caveat fallacy" (Ball, 2016). This is a similar effect to what was seen in the science journalism context, whereby the inclusion of hedges about scientific findings had little effect on what readers take away. Of course, all we can observe for now is the lack of effect, not the reasons why.

It is worth noting that two other experimental conditions – T2 (Rich Background) and T4 (Numeric CI) – had significant effects in the same direction of T3 (Expert Interception) for one vignette (the "health kits" scenario) but the effect was not significant nor consistent across the other two scenarios. The lack of consistent effect here is interesting and signals that more research is needed on the specific properties of each of these treatments as well as the contexts in which they are deployed. One initial hypothesis is that the effects of these two conditions are vignette-specific and more work is needed to identity the specific properties of the scenarios where these messages have impact. Another hypothesis is that the boundary condition between T2 “rich background” and T3 “expert interception” requires added precision and the vignette-specific versions tested here had subtle differences that need to be identified and refined (see Appendix B for the full language used for each of the three vignettes). The lack of effect of T4 is not too surprising, as previous findings about confidence intervals have been inconsistent: numeric expressions of uncertainty may lead people to more effective decision-making in some contexts (Joslyn & LeClerc, 2013; Kadane & Koehler, 2018) and to be ignored in others (Tversky & Kahneman, 1974). When ignored, one possibility is that the lack of effect may be explained by “variance neglect” (Vivalt & Coville, 2018), a type of heuristic people use when reasoning with statistics. Overall, the importance of these two message types and the inconclusive results here suggests the need to generate and experimentally test further hypotheses about these important uncertainty messages.

For now, this study confirms on the one hand the anchoring effect of numbers on human decision-making, illustrates how “expert interception” messages can add information that de-anchors the numbers to an extent, and that simple caveats have no effect.

5.2.2 Trust in Information Provider

According to Lupia (2013), source credibility is “the extent to which an audience perceives a communicator as someone whose words they would benefit from believing” (p.14051). Jensen, citing Karl Popper, argues that communicating scientific uncertainty helps maintain the trustworthiness of scientists, and found empirical evidence of this in the case of reporting science in the news. This finding was partially verified in another study, where expression of uncertainty either helped or had no effect on perceptions of trustworthiness (Retzbach & Maier, 2015). I extended this logic to the communication of data uncertainty in these vignettes. Guided by the intent to "being transparent" about data limitations, communicators may signal to audiences that they have produced and analyzed the data as experts should and are thus credible. Indeed Heinzelman & Meier (2013) suggested that any inclusion of caveats by human rights practitioners in their reports may be motivated by this imagined gain. I hypothesized that the treatments would increase the perception of trustworthiness of the information provider (H4), but I *do not* find any of these uncertainty messages relate to meaningful gains or losses in source trust.

Figure 3: Perceptions of Trust in Information Provider, Standardized to C1

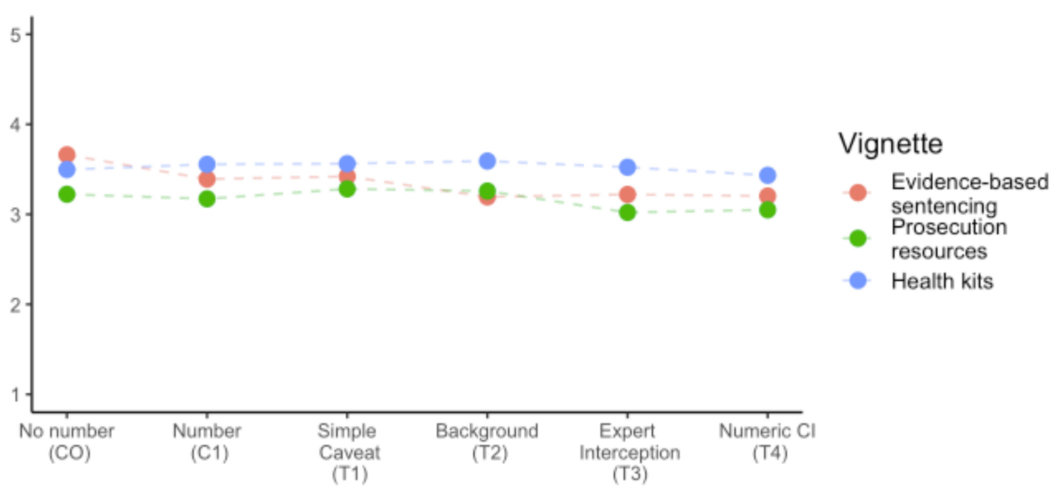
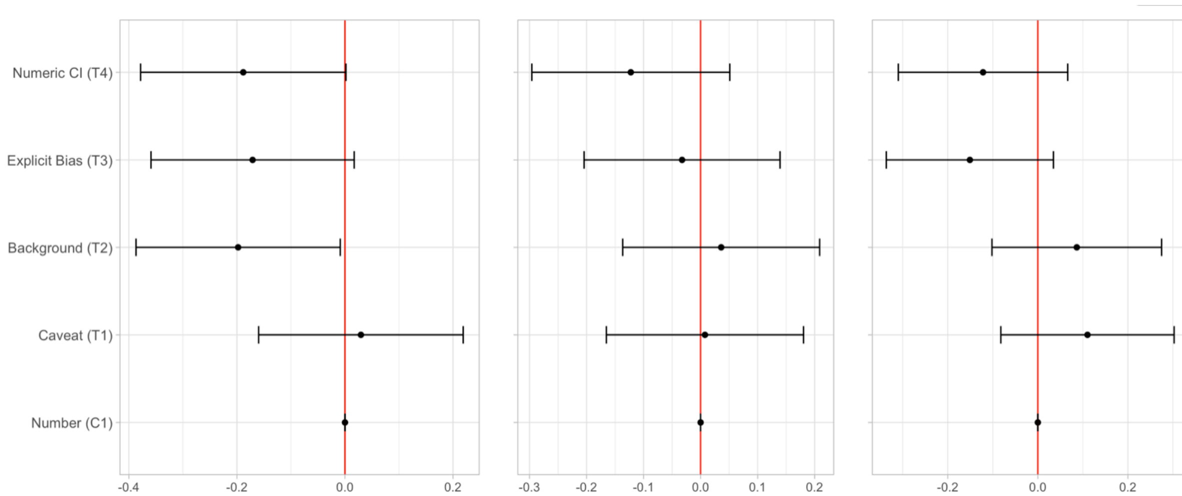


Figure 4: Perceptions of Trust in Information Provider

"Evidence-based sentencing" (left) *"Health Kits" (center)* *"Prosecution Resources" (right)*



** Vignette-specific results on Likert scale, with the point estimate and 95% confidence intervals of the effect size for the contrast between C1 and each treatment (linear regression tables below).*

Table 6: Regressions for Perceptions of Trust

C0 vs. C1:

	Evidence-based sentencing		Health kits		Prosecution Resources	
Treatments	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Number vs. No number	-0.27 (0.09)	**0.003**	0.06 (0.09)	0.534	-0.05 (0.10)	0.626
Observations	378		378		341	

C1 vs. all treatments

Predictors	EBS Estimates	EBS ctrls Estimates	SVN Estimates	SVN ctrls Estimates	PRA Estimates	PRA ctrls Estimates
(Intercept)	3.39 *** (0.07)	2.71 *** (0.13)	3.56 *** (0.06)	3.02 *** (0.13)	3.17 *** (0.07)	2.54 *** (0.14)
TreatmentSimple Caveat (T1)	0.03 (0.10)	0.05 (0.09)	0.01 (0.09)	0.01 (0.09)	0.11 (0.10)	0.10 (0.10)
TreatmentBackground (T2)	-0.20 * (0.10)	-0.15 (0.09)	0.04 (0.09)	0.06 (0.09)	0.09 (0.10)	0.05 (0.09)
TreatmentExpert Interception (T3)	-0.17 (0.10)	-0.14 (0.09)	-0.03 (0.09)	-0.01 (0.09)	-0.15 (0.09)	-0.17 (0.09)
TreatmentNumeric CI (T4)	-0.19 (0.10)	-0.16 (0.09)	-0.12 (0.09)	-0.11 (0.09)	-0.12 (0.10)	-0.14 (0.09)
ebs_often_read		0.16 *** (0.03)				
math_fam		0.10 ** (0.03)		0.06 (0.03)		0.04 (0.03)
poliv_score		-0.16 *** (0.02)		0.08 *** (0.02)		0.11 *** (0.02)
svn_often_read				0.13 *** (0.03)		
sex						0.21 *** (0.06)
pra_often_read						0.17 *** (0.03)
Observations	952	952	952	952	867	867
R^2 / adjusted R^2	0.011 / 0.007	0.101 / 0.095	0.004 / -0.000	0.047 / 0.040	0.015 / 0.010	0.083 / 0.075

• * *p<0.05* ** *p<0.01* *** *p<0.001*

Providing numbers does not increase trust at all, and in the "evidence-based sentencing" vignette, subjects rated trust slightly lower (on average) with Numbers (C1) than when No Number (C0) was provided (in this case, no risk score provided); the inclusion of the risk score lowered trust 0.3 points on the Likert scale (p-value 0.003), as seen in * *Vignette-specific results on Likert scale, with the point estimate and 95% confidence intervals of the effect size for the contrast between C1 and each treatment (linear regression tables below).*

Table 6: Regressions for Perceptions of Trust. For the other two vignettes, there appears to be a tendency for uncertainty expressions to slightly *decrease, rather than increase* source trust perceptions, although these do not cross a .05 significance level.

It is possible that the lack of findings indicates that information, quantitative or otherwise, may not be the dominant signal or most important factor for signaling trustworthiness. Indeed, name recognition, for example, has been shown to have more prominent effects in other contexts (and is what interviewees in the interview study stress as well). In these fictional vignettes, I did not account or test for these potential effects. Further, one-shot scenarios such as done here will be limited in gauging credibility. For the time being, it is only possible to say that increases to source trust due to disclosing uncertainty did not replicate.

5.3 Study Limitations

This study was an initial effort to test whether minor differences in language about data uncertainty would change mean decisions and perceptions of trust. In this first step, I have not tested the mechanisms that may be driving any observed efforts.

With respect to the observed effects, theory provided the guiding logic to the language proposed for each intervention. However, precise wording of each treatment condition could have influenced the findings, for example, the different effects seen for T2 and T4 in the "health kits" scenario but not in the others. For this reason, a next step would be to test conceptually equivalent treatments in other scenarios. This would provide further evidence about the robustness and reliability of whether the effects found here depend on the specific vignettes, or alternatively, if they hold across multiple scenarios.

Finally, asking laypersons (online M-Turk workers) to imagine themselves in the position of an advisor to an organization is an imperfect measure. This population does not represent the many contextual factors at play for many of the audiences that human rights practitioners try to inform or influence. Laypersons likely do not have the contextual knowledge and pressures of real decision-makers to influence their decisions. For example, Kahan et al. (2016) found that professionals tend to be less swayed by the motivated reasoning heuristics than laypeople. Thus, it will be ideal to conduct this study with professionals in addition to laypersons. This initial study strongly suggests that it may be worthwhile to attempt the challenging study among professional decision-makers.

5.4 Conclusion

In this study, I investigate a question at the heart of the human rights fact-finding debate: how can human rights practitioners that report numeric information prone to underreporting and bias "be transparent" about the data's limitations, and what is the potential impact of doing so.

The main results are as follows: First, I confirm that numbers do indeed serve as strong signals to influence decision outcomes consistently. Second, I find that at least among the messages I tried, only messages that make specific recommendations about how to counteract bias in the decision task consistently attenuate decisions that would otherwise be highly influenced by numbers. Third, in the first attempt to rigorously determine whether simple caveats about numbers have any impact on decision-making, I found no evidence for caveats having an effect. Fourth, I also found little evidence that perceptions of trust increase through a range of attempts at transparency when reporting uncertain numbers – in fact if anything, the results are suggestive of a trend that trust decreases with additional messages about data uncertainty.

Given that simple caveats are the closest message type to those seen used in human rights reports, these findings have important implications for practitioners. Those who produce and present information have deep knowledge about data production and could serve their audiences well by being more explicit not only about the data they do have, but also about data that is likely missing. Doing so may help their audiences comprehend and calibrate on the basis of information effectively. Given evidence that simple caveats do not appear to signal information nor credibility, their widespread use in human rights should be revisited. Finally, I believe these insights have far-reaching implications, as there is nothing inherently dependent on these results being about crime and violence data. I recommend such uncertainty messages should be tested in other contexts where numbers about hard-to-observe phenomena are in high supply and demand.

Going forward, more research should be done to test whether human rights practitioners are willing to include messages that communicate bias explicitly in their reports. Findings from the qualitative interviews suggest that this will likely be a political decision, as any benefits to conveying uncertainty will be weighed against the pragmatic and political costs of doing so. However, advocates --as well as scientists and many others – also convey data for many purposes where those concerns are not present in the same way. There may be many other situations in which advocates and others will want to convey uncertainty effectively and the results in this experimental study can support those moments. For example, in the criminal-sentencing context included in one the vignettes, the main human-rights-related concern is precisely that decision-makers will over-rely on the data in ways that magnify underlying biases. Advocates here can draw on the findings in this study to encourage improved uncertainty communication for criminal justice decisionmakers.

Chapter 6 Discussion and Conclusion

I set out to understand the viability and impact of a common petition: for human rights advocates to "transparently communicate" data uncertainty. My findings contribute to the growing scholarship taking a critical look at the use of human rights statistics.

The study is prefaced on the inherent uncertainty of human rights violations data and inquires how this fact shapes human rights information politics and data-informed decision-making. I conducted a mixed-methods study consisting of: (1) qualitative interviews with human rights advocates to understand how they think about the value and impact of making quantitative claims and how they manage the associated uncertainty when using numbers to inform and influence audiences; and (2) a complementary survey experiment to test theoretically-informed hypotheses about the impact of different ways of communicating data uncertainty to people that make decisions on the basis of the information.

As a whole, I find that human rights advocates must make trade-offs about how to manage the uncertainty. In this chapter, I first bring together my key findings from the interviews and the survey experiment, elaborating on how they relate to existing scholarly literature. I then offer a novel framework which I call the "rigor-pragmatism continuum." The continuum enables a more productive analysis of the various trade-offs faced by human rights advocates, and it also re-interprets previous debates about human rights data. I will discuss some of the political and practical implications of my research for human rights advocates and others presenting uncertain numbers to audiences. I close with directions for future research that build upon this work.

6.1 Key Findings

1. **Data uncertainty is an unavoidable reality in human rights work, and human rights advocates are keenly aware of this.**

Advocates experience a palpable tension between the high advocacy value ascribed to numbers and "data uncertainty." Advocates unabashedly confess the political utility of presenting uncertain numbers as an advocacy tool and their willingness to use it. At the same time, they have deep awareness of limitations, bias, and gaps in the data – and they contend with them, as we have seen.

While many scholars have previously found the performative utility of flawed numbers for advocacy (Andreas & Greenhill, 2010; Best, 2012; Cohen & Green, 2012; Root, 2016), I believe they too often jump to harsh conclusions. For example, "Peter Andreas and Kelly M. Greenhill see only one problem: these numbers are probably false" (book abstract), and Cohen and Green see the incentive for advocates to produce "suspect facts" (p.446). However, this line of reasoning leaves unaddressed that fact-based evidence about people killed in conflict, victims of rape, hate crimes, and other data about such hard-to-document phenomena are and will remain inherently uncertain. Furthermore, I find that such previous work has paid to little attention to how advocates themselves handle the tension between high data demand and uncertain data supply. Through my interviews, I find that the advocates interviewed do not simply ignore or discount data problems, publishing any big, attentiongrabbing figure available. Instead, they make highly pragmatic, yet principled choices and trade-offs about how they ultimately use uncertain numbers in their communication strategies. Those trade-offs, which I find to be highly context-specific and elaborate further below, involve time, resources, expertise, and rigor, among other things.

2. **Human rights advocates are generally conscientious about wanting to avoid "suspect facts," and one key way they do this is by making use of good enough numbers.**

While a few organizations in this study strive to use rigorous statistical methods to produce robust and reliable numbers, most of the advocates I spoke with expressed consistent and shared ideas about more pragmatic, yet principled alternatives. For example, they hedge their numeric reports with phrases such as "at least" or "no more than" – claims they can make with high confidence. Akin to the notion of "satisficing" (Simon, 1956), I call these *good enough numbers* and argue the ideas underlying them constitute a set of pragmatic ethical and political norms shaping human rights information politics in the advocacy space. While advocates know good enough numbers to be uncertain, they insist that they are temporary, conservative, and empirically-based – not exaggerated or unfounded guesses. As such, advocates differentiate these claims as "snapshots," not robust "truth-claims." Interestingly, while transparency about data and methods is discussed as an ideal for good enough numbers, it mostly exists as a tradeoff. Language about methodology, data provenance, and data limitations is often minimal, as discussed in detail below.

As discussed in Chapter 2, Porter's definition of objectivity is specifically about "distance from subjectivity" which is achieved through a systematic process of data production. He emphatically distinguishes objectivity from ideas about "truth in nature." I find that human rights advocates intuitively share this conception when they reason about using good enough numbers. They consciously use the performative power of numbers, even where the numbers are not certain. They are comfortable that hedged presentations of numbers are ethical and valid in the short-term and come with minimal cost.

These findings are also worth considering alongside Sally Merry's work (2016) about the "seductions of quantification" of human rights indicators. Arguably, there are some important similarities in the value of quantification for advocates included in this study as Merry finds among policymakers included in her study, not least that the seduction mostly relies on the suppression of uncertainty. However, the seduction differs in this context in one important way. In her study, very much in accordance with Porter's *Trust in Numbers*, Merry concludes that "[quantitative] indicators offer a particularly reliable form of truth" (p.26). Again, in this conception, quantification as a technology of distance is meant to project technocratic truth, cleansed of the subjectivity associated with politics. In the advocacy space, however, the politics is very much front and center – even with numbers. Among my advocate participants, most were conscious that they were using numbers less to make robust and reliable 'truth-claims" and more for an immediate goal of getting attention.

3. **Human rights advocates consider political and cognitive risks of transparency. Most calculate that minimal transparency aligns with audiences' expectations, thus poses little risk to them.**

In my interviews, advocates portray transparency – of methods and data uncertainty – as a frustrated ideal. Advocates and human rights scholars both underscore its value for credibility

(Ball & Price, 2019; Cohen & Green, 2012; Satterthwaite & Simeone, 2016). Yet, their incentives run counter to elaborate transparency. Most interviewees believe their audiences are not interested in methodological elaboration. Results from my survey experiment largely confirm these intuitions. Only one uncertainty message – expert interception (T3) – had any impact on the decisions subjects made with numbers. Furthermore, there was minimal fluctuation in the trust measure across treatments and controls with any of the different uncertainty messages. Overall, while methodological transparency may be an ideal, transparency does not appear to be a game changer with respect to trust in the source.

Some participants even feared that too much transparency could be politically risky and cognitively complex, aligning with Merry's finding in global governance forums, that "[i]n order for an indicator to succeed in policy and public domains, it must present information in a simple and unambiguous way without a great deal of qualification and methodological discussion" (p.20). In light of these findings, it was not surprising to find little reference to data limitations among the twelve randomly selected Human Rights Watch reports that I analyzed. In human rights work, robust transparency is not sought after, and there is little cost to avoiding it.

While skeptics may interpret the ideas underlying low transparency to be self-serving for advocates, they very much echo previous findings from across science communication domains. In politicized contexts like climate change and tobacco policy, Oreskes & Conway (2010) find high political risks to conveying scientific uncertainty. Similar to the idea of "seduction," Manski finds a "lure of incredible certitude' for most scientific claims made in public policy forums more generally (2018). Decades of research suggests that when presenting science in politicized contexts, it is most "effective" to maximize audiences limited attention, leaving little room for methodological discussion and caveats (Lupia, 2013).

My experimental findings also support and extend the insights from cognitive research. As reviewed earlier, studies show that some specific kinds of presentations can ease the effortful thinking required to bypass cognitive heuristics, thereby supporting "more effective" decision-making with uncertainty (Fernandes et al., 2018; Joslyn & LeClerc, 2013; Kadane & Koehler, 2018; K. A. Martire et al., 2013). The one message that moved decision-making away from an anchoring number was explicit about the impact of the bias on the decision task. This result replicates the findings from Joslyn and LeClerc (2013), adding weight to the evidence that "preferred deterministic interpretations" do affect information consumers.

Less direct treatments (i.e., a simple caveat expressing the data were unrepresentative (T1), rich background information about the data production and its biases (T2), and numerical confidence intervals indicating the direction and degree of the data bias (T4)) did not have consistently significant effects on shifting decision means away from the numerical anchor (as compared to control across three vignettes). Of note, across all three vignettes, differences between decisions made in the "simple caveat" condition versus those made with a "number, no uncertainty" were consistently very small, the weakest at attenuating the anchoring power of the number. By failing to reject the null hypothesis of C1=T1, the experimental results leave intact the idea of the "caveat fallacy" (Ball, 2016) – i.e., that simply warning people that a number may be unrepresentative causes little change to the ensuing decision making. This finding is also consistent with Jenson (2008), who found that hedges have little effect on what readers understand about broader findings presented in the science journalism context.

4. **"Expert interception" – the message that most consistently moves people away from a biased, anchoring number – intentionally asks audiences to accept data producers' expert opinion about data limitations instead of mechanically placing their trust in the numbers.**

A key feature of the "expert interception" message (T3) is that it conveys the uncertainty as it specifically relates to the imminent decision. To do so, it explicitly draws on the data producer's *expert* knowledge about the shortcomings in the data and uses that knowledge to anticipate and warn the decision-maker. In their discussion about why such a message type was "effective," Joselyn and LeClerc emphasize its context-specificity, but I believe we gain added insights from Porter. Recall that Porter finds quantification to be a valuable communication tool when trust in the subjective authority of an information provider is weak. Perhaps a corollary to this finding is that numbers exert less power when subjective authority is strongly projected to counter the number.

In the T3 condition, the information provider is essentially communicating a subjective yet expert signal to audiences that they should err on the side of trusting their expertise about the data limitations instead of heuristically trusting and mechanically accepting the numbers as valid. In this way, both the number and the expertise of the information provider exert authority, and the decision-maker bears the onus of whether they should heavily rely on the number in their decision or trust the expert's interceptive message.

One could imagine that a choice to formulate the message in this way depends on how much the information provider believes the audiences is receptive to a heavy expert hand. In the human rights advocacy domain, one could imagine that such an explicit message aiming to shift trust in numbers toward trust in experts may undermine the rhetorical appeal of simple and unambiguous numbers.

5. At least in human rights advocacy contexts, trust in institutional authority can matter more than quantitative expertise.

As seen in the qualitative interviews, I find that the perception of the institutional source authority is cited as the key heuristic advocates use to evaluate and cite numbers produced by others. Advocates admit they rarely verify the quantitative expertise behind numbers they are willing to cite if they recognize the source is offering good enough numbers. In these cases, avoiding reputational risk when citing numbers' produced by others sometimes means making political calculations about citing "official sources" and other times means relying on trusted institutions in human rights community. In such cases, the risk is assumed to rest primarily with the authoritative source being cited.

Porter explains that trust in numbers is a powerful force precisely because they communicate complex information in a simplified form. As audiences engage in such trust, they assume the numbers were produced according to expert, rule-based quantification procedures. But the unambiguous presentation of numbers itself does not convey that the assumption of rule-following was met. To avoid having to deal with the verification process (which would be difficult to conduct anyway given how little methodological information is said to accompany most numbers), advocates shift to citing authoritative sources that for the various reasons presented above, absorb the perceived potential risk to the citers' reputation.

6. **Advocates face a clear choice to be made between allowing numbers to speak for themselves or being proactive about mitigating bias.**

Based on the evidence from the experimental survey, which demonstrates a substantive and consistent difference between C1 (number-only control) and T3 (expert interception), it becomes clear that information providers face a clear choice about whether to use an uncertain number and about the extent to which they will support the audience to understand the bias and limitations in a number. Because the "expert interception" message seeks to correct for the ways the data bias may impact decision-making, human rights communicators are *necessarily* making a choice between letting the number speak for itself, and intentionally correcting it. There is no technocratic truth, especially under uncertainty. So if advocates, or any information provider, believe it is important for audiences not to interpret a number with a false sense of precision, they must exert their data expertise and convey the bias explicitly.

6.2 A Rigor-Pragmatism Continuum

The findings above suggest we need a new model to analyze the practical trade-offs that advocates face when they consider using uncertain data to advance their work – a model that moves beyond the dichotomous good/bad framing. I propose a continuum, with rigor at one end and pragmatism at the other. Rigor refers to the effort made to adhere to the rules of a system (Marquart, 2017), and in this context, the effort made to reach and express the full, unvarnished truth as far as it can be known. Pragmatism, as a philosophy, is an approach that assesses the validity of specific ideas by the consequences of their practical outcomes (Peirce, 1966). On such a continuum, anyone can undertake choices and gauge potential costs that pull them towards rigor on one end and pragmatism in the opposite direction.

A valuable quality of the rigor-pragmatism framework is that the full spectrum has validity *in context*. Rigor is valid for its stress on truth; pragmatism is valid for its emphasis on utility and action. This reframes the prevailing debates about numbers in human rights, which tend to judge them as "good" or "bad" (Best 2012; Andreas & Greenhill, 2010). The rigor-pragmatism continuum is better-suited to understanding and interpreting human rights numbers, because very few in the community put out outright "bad" numbers – instead, my participants regularly and conscientiously wrestled with how and when to use less-than-perfect numbers. This model can be extended and enhanced in future work.

Some of the ideas inherent to the rigor-pragmatism continuum have similarities to the points Sheila Jasaonoff offers the idea of "serviceable truths," which refer to scenarios where science is meant to support reasoned decision-making in policy spheres (2014). This often requires understanding that the standards to which science-inflected claims are held are more practical than they would be in pure scientific communities that depend on the "sharp critical gaze" of methodological peers (p.1738). She urges scholars that study science-in-policy domains to go beyond analyzing science *in action* (focused on analyzing how the facts get made) to analyzing science *for action* – (focused on explaining how science can support policy goals).

I believe that in the human rights advocacy realm, we see a similar "serviceable" standard in operation.[89] The rigor-pragmatism continuum that I propose here paves the way for a more nuanced analysis of how advocates pursue satisfactory alternatives to using uncertain data that go beyond conducting rigorous statistical analyses or abandoning the use of such data altogether (what I called the "fix the data" or "drop the data" in Chapter 2: Literature and Theoretical Framework).

[89] In the human rights scholarship, there is a larger debate about human rights practice becoming ever-more pragmatic (see Dancy, 2016; Sharp, 2018). In their increasing use of quantification – and especially highly uncertain numbers – this study offers a data point in favor of some highly pragmatic practices by human rights advocates.

Rigor --- Pragmatism

Table 7: Data uncertainty forces trade-offs on a rigor-pragmatism continuum for many issues involved in producing and presenting numbers

Issue on Continuum	Rigor	Pragmatism
Goal advanced by numbers	Truth claims	Rhetorical claims
Temporal robustness of numerical claim	Enduring	Temporary
Reliability of numbers	Higher	Lower
Context applicability	More contexts	Context-specific
Time available to produce numbers	Slower	Faster
Resource required for analysis	Higher	Lower
Quantitative expertise required to produce numbers	Higher	Lower
Perceived risk of push-back about data uncertainty	Higher	Lower
Who bears onus of uncertainty	Communicator	Others
Transparency about data uncertainty	Elaborate	Vague/concise/absent

I have included in the table a range of topics that advocates brought up during interviews that they consider as they use uncertain data for advocacy. The first issue refers to the goal sought – whereby advocates differentiated between pursing "truth" claims with science to pursing more persuasive end by drawing on numbers in a performative way. Temporal robustness distinguishes between claims meant to endure over time versus claims that are deliberately meant to be "snapshots," as many advocates expressed. Reliability of numbers refer to the extent to one can arrive at the same results when repeating the procedure that produced the result, which may rely on stable and knowable data and procedures. Context applicability is an especially important issue in the table, as it refers to the degree to which the numeric claim is valid in a specific context versus to its validity across contexts (I return to this in more detail below). Time, resources and quantitative expertise all involve degrees of effort and material costs. Anticipated push-back refers to the extent to which communicators foresee and prepare for methodological critique from potential audiences. Who bears the onus of uncertainty refers to the extent to which communicators view themselves as liable for the uncertainty inherent to the claims (and opt for rigorous methods to contend with it), versus the extent to which they shift that responsibility either to the institutional sources they cite or to the audiences that may have to deal with any interpretation and use of uncertain numbers. Finally, while transparency is often coupled with rigor, this study shows how it is also better understood on a spectrum, i.e., one can have more or less of it.

My experiment also tested different versions of transparency implementations which could be gauged along the same spectrum. Again, a key finding here is that the degree and form of transparent communication of uncertainty will be context-specific and clearly involves trade-offs with some of the other items in this table. "Total" transparency can be considered more rigorous while pragmatism might dictate conciseness or vagueness.

Among the strengths and reasoning I heard from some interviewees for opting for more rigor is that numeric claims about human rights violations should "speak truth to power." As such, they should be more enduring and robust across contexts. As a result, they will require relatively higher effort (expertise, time, resources) to produce. When a communicator opts for rigor, perhaps by privileging in-house statistical expertise or out-sourcing such work, they also accept greater responsibility and thus accept accountability about the impact of the uncertainty on the claims they make. Groups that opt for this level of effort often limit the contexts for which they do so to those they consider "high-stakes," such as when they want the data to inform policy, when they anticipate resistance from adversaries (political, legal, or public opinion), or when they aim to make more long-lasting empirical contributions (e.g., truth commissions). Only a few organizations represented in this study aim towards this level of rigor in their production and presentation of uncertain numbers.

At the same time, rigor absolutists should also recognize the trade-offs that pull towards pragmatic options about how to contend with data uncertainty. By far, the majority of human rights advocates in this study make choices that pull toward the pragmatic end of this spectrum, doing what they can for rigor, but making concessions so as to achieve advocacy goals within the time and resources they have. As these advocates think about using the rhetorical power of "good enough" numbers, they choose to expend relatively low effort and resources to transparently communicate uncertainty.

Theoretically, pulling towards pragmatism in the use of quantification as a communication tool is consistent with Porter. He finds much idiosyncrasy and pragmatism in the production and presentation of numbers in very different contexts. However, among the highest tradeoff I see in good enough numbers is their ephemeral validity, as "snapshots" in time. Indeed, the idea of temporary numbers defies the nature and power of numbers – not only do they have longevity and fluidity across contexts, but among the greatest concerns with their misuse is that they are taken out of contexts (Best, 2012).

Advocates mostly see near-term attention goals over long-term use and reuse of numbers. However, advocates violate these assumptions all the time. They themselves may reuse an existing number not as a temporary claim (for example, about the known number of victims to date) but as a reference point to claim that violence is escalating, the context-specificity of the original number is breached. When making such claims about changes over time or other patterns of interest (which we saw is another of the main goals for advocates), the initial number becomes a hard data point used for comparing time past with time present. A context violation in this way can occur by a single author using information in different ways in two different moments (e.g., a cautious claim of magnitude in moment one and as a hard reference frame in an over-time comparison in moment two). Another kind of context violation with uncertain numbers is while they may have low cost in general advocacy contexts, they easily become consequential anchors in subsequent decision-making forum (such as the scenarios included as vignettes in the survey experiment). Therefore, the short-term utility of "good enough" numbers contrasts with their long-term staying and anchoring power. As a result, future users may bear the onus of not being aware of or being able to warn users of data uncertainty because it fell away in earlier moments of publication.

Andreas and Greenhill (2010) close their book with the following: "Yet however much one may try to rationalize the political uses of bad data, a failure to at least strive for, if not achieve, statistical accuracy and honesty tends to be not only unhelpful, but can deepen public cynicism and distrust and erode credibility" (p. 272). While, as is clear by now, I disagree with the "bad" data characterization, I nonetheless agree that striving for the rigorous end of the spectrum is worthwhile. Rigor should matter not least because of the influential anchoring effects of numbers, but also because the idea of narrow-conceived numbers for advocacy contexts is unrealistic. To the extent possible, investing in rigorous practices – including investing in well-established ways to make data-based inferences and being transparent about data limitations – is worthwhile. While it will always be impossible to reach, rigor can be a North Star pulling the community in more robust and transparent directions.

6.3 Refining Appeals to Rigor, Transparency, and Credibility

As we have seen, in recent years human rights scholars have cast a critical eye to the use of numbers in human rights advocacy. Almost unanimously, they emphasize that advocates should strive to be more rigorous and more transparent. Beyond the findings and continuum offered here, this research sheds light on the limitations of existing appeals and suggests ways to improve them.

At present, appeals to rigor and transparency take at least three forms. Some scholars appeal to the interest of organizational credibility (Cohen & Green, 2012; Satterthwaite & Simeone, 2016). For example, Satterthwaite and Simone write: "Greater transparency and rigor will ultimately enhance [advocates'] credibility with target audiences" (p.322). While my study does not discredit such ideas, per se, it also does not find that opacity necessarily leads to loss of credibility. In the experiment, we see that trust in sources does not appear to be highly associated with the presentation of numbers and uncertainty. In the interviews, advocates mostly did not raise credibility penalties in any examples they provide on the use of uncertain numbers. They hint at judging credibility and gauging trust in numbers based on other metrics (e.g. political position or institutional standing of the source), although further research is needed to understand what generates institutional trust in the human rights community, including laypersons trust in institutions and trust across peer organizations within the community. Meanwhile, advocates make pragmatic trade-offs to protect concerns about source trust and credibility.

Others scholars makes appeals to rigor by citing that inaccurate numeric claims can lead to dire potential consequences for truth and accountability (Andreas & Greenhill, 2010; Ball & Price, 2019). However, such appeals for more rigorous methods to better approximate "truth" may fail to resonate with advocates that believe "good enough" numbers – numbers based on data that they believe to be and clearly state to be lower bounds – suffice for near-term advocacy. What becomes clear in the interviews is that truth-seeking and activists goals are important, yet different; data uncertainty forces a trade-off about what to seek in a given context. Therefore, while advocates in this study would likely agree with appeals to rigor for truth-seeking, such appeals may ultimately be weak motivators to change advocacy communication practices.

Finally, to motivate critical information consumption, some of the existing literature either explicitly or implicitly characterizes uncertain data as "bad" data and the lack of rigor and transparency in the use of it as "misuse." They leave a subtext portraying information providers as suspect a priori. Sometimes scholars must include awkward clarifications to suggest that is not necessarily what they are implying.[90] Such appeals would serve information providers and

[90] For example, Cohen and Green (2012) add, "To be clear, we do not argue that NGOs intentionally use or create inflated statistics" (p.446).

information consumers better if in their appeals, they recognized the contentious trade-offs involved in information presentation, where good enough numbers with low transparency is largely seen to be working.

While valuing the complex tradeoffs that advocates make to advance their agendas, my appeal is for advocates to more seriously contend with the fact that numbers will not be constricted to the temporary and context-specific bounds within which they are originally published. The short-term utility of "good enough" numbers contrasts with the unavoidable fluidity, anchoring effect, long-term staying power and reuse of numbers. For this primary reason, I believe advocates could consider conveying data provenance and limitations to a greater extent than they currently do. At least in the experiment included here, we see little penalties to trust in the information provider when including any of the uncertainty messages.

Ultimately advocates and any other information provider about human rights abuses face a choice about how much to communicate data uncertainty underlying the numbers they publish to near-term *and long-term users*. The choice may be hard. In the near-term, advocates may reason that they have more to lose than to gain by transparently communicating methods and data uncertainty (as many interviewees expressed). But it is worth pausing for a moment and considering future users. Would the interpretation of this number as a hard anchor (which is likely) by future users be problematic? If so, can they design expert interception language to mitigate the risk to future users? Or can they convey uncertainty with rich background language for posterity?

In the process of investigating the challenges and possibilities of conveying uncertainty in advocacy and decision-making scenarios, this research opens many more questions. Hopefully future research can help us understand the specific impact of using language to communicate uncertainty in real-world contexts that brings together the political and cognitive insights, as well as the challenge of uncertainty communication that serves near and long-term users. Considering potential future users of uncertain information is important for several reasons. For one, advocates *know* the numeric information to have an uncertain relationship to the issues they represent. It is important because numbers endure over time and across contexts. It is important because numbers have an outsized impact on their readers whether to impel attention (as seen in interviews) or to anchor decisions (as seen in the experiment). And, it is important because here the numbers in question are about human rights atrocities – information with an expansive and diverse set of stakeholders (ranging from immediate victims and families, to future generations seeking answers about what happened and who did what to whom). In sum, my appeal is to include information about data uncertainty for posterity.

6.4 Study Limitations and Future Research

The survey experiment on communicating data uncertainty conducted here is the first of its kind in the human rights literature. It was useful to begin by conducting an initial study that tested whether the conceptual interventions had any significant impacts on outcome measures among laypersons. To do so, I focused in this experiment on designing vignettes and interventions (treatments) that were consistent with scholarly insights from science communication and human rights scholarship and practice. While I used previous scholarship to guide my logic to the language proposed for the interventions, precise wording of each treatment condition may have influenced the answers provided. For this reason, I included multiple vignettes covering different human rights contexts where quantitative data is used to inform or influence the information consumer. This enabled me to detect whether the effect of a given presentation of data uncertainty treatment depends on a given vignette, or alternatively, if it is relatively stable across vignettes.

Nonetheless, the experiment was limited in several ways. First, asking laypersons (online M-Turk workers) to imagine themselves in the position of professional decision-makers is an imperfect measure. This population does not represent the many contextual factors at play for many of the audiences that human rights practitioners try to inform or influence. Laypersons likely do not have the contextual knowledge and pressures of real decision-makers to influence their decisions. Research from other domains suggest some experts are less affected by framing effects (Gächter, Orzen, Renner, & Starmer, 2009), and that professionals are less swayed by motivated reasoning (Kahan et al., 2016). Going forward, it will be ideal to conduct this study with professionals in addition to laypersons, although recruiting a sufficient sample size with internal validity could be challenging. Ideally such a study could test not only how uncertainty messages are received (as tested in this dissertation), but also how willing real-world practitioners are to send such messages.

Also, while we gain evidence of that "expert interception" (T3) was the uncertainty message most "effective" at shifting decision means consistently and significantly, we need more work to experimentally isolate *why* treatments did or did not have an effect, especially why two of the experimental conditions produced different results across vignettes. This could be done with additional hypotheses and round of iterative experiments, as detailed in Chapter 5. For example, for T2 ("rich background"), we do not yet know if this treatment was ignored altogether, or was considered and discarded as information that merited a different decision from what it would have been in the absence of this information.

A future study would also further benefit from some of the insights from the qualitative interview study to a greater degree. For one, it could try to more carefully distinguish “simple caveats” from “estimative” language – like “at least” – which is what advocates say they attempt to communicate and what I found was most used in the small coding exercise of HRW reports. Also, given that advocates believe one of the greatest potential costs to communicating uncertainty is reduced perception of them as an authority, a future study could refine and disaggregate measures for such impacts. Finally, a future study could also aim to test which messages best support reusing numbers in different and future contexts, while simultaneously minimizing losses to source trust in the present. For example, such a study could test whether a message type like T2 “rich background” would be adopted by a presenter and support a reuser in a different time and space than that which the number was presented for in the first place.

The main limitation of the qualitative research is that it was based on a small convenience sample of key informants and thus cannot represent the community’s views as a whole. As Creswell (2013) points outs, interviews gave me as researcher the advantage of control over the line of questioning to purse and dive in depth as appropriate for each interviewee. Nonetheless, the quality of data collected may be varied. Given that I already had a professional relationship with some of the people I recruited, I may have gotten a different level of disclosure from some people compared to others. Interviews are always shaped by different levels of trust and comfort interviewees may have felt with me as researcher. I am also cognizant of possible social desirability effects, expectancy effects, the inaccuracy of self-reporting, degradation of memory over time, and the deference effect.

A larger and more representative sample would enable more reliable results. Nonetheless, for now the qualitative study here offered the insights that I used to propose the rigor-pragmatism as a kind of prototype model that I hope will be useful as an analytic framework. To test the utility and robustness of the continuum as an analytic framework, a more diverse study would be valuable, expanded along many other important factors influencing HR advocates – other geographies, different levels of institutional hierarchy within organizations, more diversity in the resources of organizations, and different levels of adversity the organizations perceive from their audiences – especially because these are key trade-offs in the continuum. Finally, the coding exercise of HRW reports is preliminary and thus limited. A more comprehensive study could use the coding model developed here, but include a larger sample of reports from more organizations and more periods of time.

6.5 Conclusion

With this research, I have shed light on the political, behavioral, and methodological challenges that advocates face as they collect, communicate, and deploy violence statistics in global and local human rights advocacy contexts. Informed by the results of this research, the rigor-pragmatism continuum hopes to serve as a more nuanced analytic framework than what good versus bad framings support about how human rights advocates wrestle with using uncertain numbers.

As we appreciate the tradeoffs advocates must make with using uncertain numbers in context, we can also see some weaknesses with such communication strategies. Results from the survey experiment make clear that numbers are not read in the way advocates believe they are presenting them. The notion that numbers are temporary and performative statements contrast with their numeric anchoring power and their endurance long past immediate advocacy contexts. The experiment, on the one hand, confirms advocates' intuition that most messages about data limitations have little impact on outcomes. On the other hand, it offers evidence that with one specific message formulation, what I have called "expert interception," communicators of information can intervene and attenuate the interpretation of numbers, *if they so choose*. Perhaps easing concerns about risks to the source credibility, the experiment finds that trust in the information providers is largely unaffected by any of the forms of conveying data uncertainty.

However, such messaging requires communicators to take the position of data experts and use their knowledge to inform readers of how data limitations can bias outcomes. Intercepting the power of numbers to project "mechanical objectivity" with expert knowledge about the data generation process and data limitations could be politically risky in domains where experts are seen as suspect, but arguably there are many cases where these findings can help convey data uncertainty, for human rights contexts and much beyond (e.g., climate science and public health). Whether information providers choose to include such messages will ultimately depend on if they believe it is important to mitigate bias by future users, requiring a shift for advocates usually seeing data in the immediate term as an advocacy tool.

Appendix A: Interview Protocol for Qualitative Study

General questions about their work

- Tell me about your current position and area of expertise within the organization.
  - How did you come to your current position? (including: education, previous work experience, etc.)

- Data production: Can you tell me about the fact-finding and documentation work at the organization?

- How are you involved in that work?
  - Probe on the data generating processes.

- Tell me more about your role in producing information about human rights violations – documenting, monitoring, collecting
  - Challenges?

- Do you have experience actually conducting the quantitative data analysis by your organization?
  - If relevant, please describe a recent example.
  - How is this information used by you or your organization?

Presenting data and quantitative claims in your work

- Do you ever choose to incorporate numbers, patterns, quant arguments, data viz into your reports? Conversations? Advocacy in general?
  - Do you ever choose not to?

- Are there specific audiences for whom you believe this appeal works? Any this appeal isn't useful for? What other "types" of appeals for other audiences? (emotional, informational, stories, others?)

- What do you perceive to be advantages in using quantitative evidence (metrics, numbers, aggregate data patterns) in messaging?

- When do precise estimates with known error ranges matter? Examples?

- What you think presenting quantitative information uniquely helps you achieve?

- Do you have any examples of times when you think statistics or quantitative claims were instrumental in achieving some intended outcome? Elaborate. Why was it successful?

- Do you have any examples of any specific decision-making scenarios specifically influenced or able to influence with numbers? Something analogous to EBS?

- Do you have any examples of times when you think the inclusion of statistics or quantitative claims in a publication or argument resulted in a problematic outcome? Elaborate.
    - What may have been a more favorable outcome?
    - Why?

- Have you made any changes to how you (or your organization) presents quantitative data to inform and influence agendas over the past 5 years?
    - Why do you think it's changed in this way?
    - Have those changes been successful?
    - Are there any weaknesses or drawbacks that have resulted from those changes?

- What do you perceive to be disadvantages in using quantitative evidence (metrics, numbers, aggregate data patterns) in messaging?

- Do you ever find the advantages and disadvantages in tension with each other? Explain.
    - Do you ever find that the advantages outweigh the disadvantages? Explain.

- Do you ever think about the robustness or quality of the data underlying the quantitative messages you present? Data Limitations?
    - Have you ever had any concerns about presenting quantitative claims?

- Do you ever include quantitative claims or arguments in your reports that were produced by a third-party (not produced by your organization)? Explain.
    - How do you select which metrics or findings to include?
    - How do you judge their credibility?

Receiving and consuming quantitative claims:

- Can you tell me about other sources where you see or from where you "receive" quantitative information about human rights violation from? I'm referring to processed information in the form of quantitative conclusions or claims from other sources.

- How do you assess the credibility of numerical claims and trends in human rights violations?
    - Probe: Do you ask/know author? About language? Do you gauge against what you already know about a given context?
    - Do you try to gauge "how hard they've worked"?
    - How might your assessment of credibility differ across contexts where you are more or less familiar with issue/country/other contexts?
- Have you ever seen quantitative data in reports that you felt was weak or flawed? Explain.

- Have you seen any changes in how others present quantitative data to inform and influence agendas over the past 5 years?
    - Why do you think it's changed in this way?
    - Have those changes been successful?
    - Are there any weaknesses or drawbacks that have resulted from those changes?

Probes about presenting data & uncertainty in their work

- What do you usually present in a methods section of a report?
    - What do you hope that communicates?
    - When numeric data included in reports, statements, etc. – do you usually present how that data was collected? Any data limitations?

- Are you aware of any limitations or weakness in the data you collect and/or present to your audience?
    - Are those limitations/weaknesses communicated to the audience?
    - Do you believe they should be?
    - Are there differences in how you present weaknesses to your different audiences?
    - Is there any specific terminology you often employ when communicating data weakness?

- How does your presentation of information and its limitations differ between audiences, such as public and private reports?
    - Between potential adversaries and potential allies?

- Has your expression of the origins of the data, or any uncertainty about what it includes or excludes ever posed any problems for your work?
    - If relevant, please describe a recent example.

- In the literature, call for more transparency about data production. What do you understand this to mean?
    - What do you understand a caveat to be?
    - Do you use caveats when you present quantitative information?
    - To express what?

    - How do you perceive caveats when you read them in reports by others?

- Any downsides to presenting numbers in short and long-term?

Appendix B. Vignettes and Treatment Language

Vignette 1: Health Kits for Wartime Sexual Violence

Stop the Violence Now (SVN) is a non-profit agency dedicated to promoting human rights globally. We are headquartered in New York, but collaborate with local authorities around the world to eradicate violence and its effects.

We are concerned about the long-term health of rape survivors. Studies show that survivors of sexual violence are more likely to experience mental health issues such as depression, anxiety, substance abuse, and post-traumatic stress disorder. They also have higher rates of illnesses like cancer, cardio-vascular disease, and diabetes over their lifetimes.

We have decided to provide post-rape care resources to victims of sexual violence in war-torn countries. These come in the form of "kits" – which include a bundle of professional services promoting physical and mental health, and social services.

We will begin our provision of these kits in Colombia. This year, SVN's offices documented increasing numbers of sexual violence in two distinct locations in the country. One is the large and remote Pacific region of Choco where there is still an armed conflict. The other is a displaced persons camp in the northern region of Sucre. We will distribute emergency health kits in these two regions.

[* C1+ Treatment]

We aim to distribute our limited services to benefit as many victims of sexual violence as possible who arrive at the locations where we offer services. Often, we find that more people show up for services than official counts have tallied.

We currently have the resources to fund 700 kits. We are seeking input on how to best distribute them.

Based on the information provided here, please tell us how many of the 700 kits you would allocate to SVN's office in Choco: _______

Table 8: Treatment language Vignette 1

Treatment	Name	Uncertainty message [*]
C0	No numbers	[No C1 paragraph above]
C1	Number, NO uncertainty	In Choco we have 200 documented cases of sexual violence and in the Sucre displaced persons camp we have 400 documented cases.
C1+ T1	Simple Caveat	Be aware that reporting sexual violence is culturally stigmatized in Colombia. Many people may choose not to report their experiences for fear of social consequences.
C1+ T2	Rich Background	In both regions of Colombia, people may avoid reporting sexual violence due to cultural stigma. In Choco, local government-run health clinics are doing their best to report known cases, yet there are very few of these clinics across the large region. More generally, data collection is challenging in a context where illegal armed groups continue to operate and use sexual violence as one way to repress the population. In Sucre, all the reported cases are from a relatively small displaced persons camp. There are several health clinics throughout the camp where people can get aid, and local officials make an effort to document all the sexual violence cases they learn about.
C1+ T3	Expert Interception	Be aware that the reported cases of sexual violence are an undercount, and likely do not represent the total number nor the pattern of all sexual violence cases in both locations. While we cannot be sure, we may know about many more of the victims of sexual violence in the Sucre camp than we do about all victims in Choco. In Choco, there are probably many more victims of sexual violence than are reported because the region is so large, precarious, and still has active armed groups. In Sucre, there may also be more victims than the reported number, yet the displaced persons camp is smaller, has administrators working around the clock, and is not located in an active conflict zone.
C1+ T4	Numeric uncertainty, conference interval	[C1 embedded] In Choco we have 200 documented cases of sexual violence, but the total number could range to be anywhere between these 200 to 2,000 cases. In the Sucre displaced persons camp we have 400 documented cases, but the number of people who have suffered some form of sexual violence could be as high as 600.

Vignette 2: Prosecution Resources

You are the Assistant to the Prosecutor at the Special War Crimes Court for the Northern African Republic. You need to advise the prosecutor on which war crimes cases to prioritize for investigation.

The Northern African Republic had an outbreak of extreme ethnic violence last year in a conflict between the two ethnic groups – the ethnic majority Borano and the ethnic minority Kaya. The international community is concerned that Borano militias may have committed genocide against the Kaya minority.

You are reviewing the evidence submitted to the Prosecutor's Office by non-governmental organizations. Most of the evidence submitted comes from one organization, the African Peace, Empowerment and Rights Group (APERG). Last year, APERG led a fact-finding mission to the North African Republic for 2 months, during which they produced an extensive report detailing the human rights violations they observed. *In that report, APERG presents the names of civilians killed, which include people from both Kaya and Borano ethnic groups.

The Office of the Prosecutor has over 200 cases on file to investigate for violent events carried out by alleged perpetrators from both ethnic groups. However, they only have the resources to review 80 cases in the foreseeable future. The cases require similar levels of effort and resources to prosecute.

The Special Court always seeks to allocate their limited resources fairly, for example, by prioritizing cases in proportion to the level of victimization by ethnic group.

In your job as Assistant to the Prosecutor, based on the information provided, consider how would you allocate the 80 cases across perpetrators targeting victims from each ethnic group.

Of the 80 cases, type the number you would allocate to investigate cases involving victims of Borano ethnicity: _________

Table 9: Treatment language Vignette 2

Treatment	Name	Uncertainty message [*]
C0	No numbers	In that report, APERG presents the names of civilians killed, which include people from both Kaya and Borano ethnic groups.
C1	Number, NO uncertainty	In that report, APERG presents the names of just over 3,600 Kaya civilians killed by Borano militias. These same investigators also documented 400 killings of Borano civilians by Kaya guerillas.
C1+T1	Simple Caveat	Be aware that APERG investigators could only document victims in areas where they had access, so all the figures are prone to underreporting.
C1+T2	Rich Background	APERG's team collected their data during seven weeks of investigation in November and December of 2016, during a period of high violence. Their research team consisted of two foreign researchers, a British man and a Kenyan woman. They conducted interviews with hundreds of witnesses and victims of violence with help from local translators. APERG had unprecedented access to interview people in the Southeast of the country, a region controlled by the Kaya (minority) armed opposition group. They also interviewed a few Borano government officials. They kept detailed interview notes and research logs, which became the basis of their report.
C1+T3	Expert Interception	It is likely that the evidence APERG collected does not represent the total number of victims nor the true pattern of civilian killings. The evidence they present in the report is based on the researchers' available time, resources, and the level of access they were granted during their field work. Since they had the most access to a Kaya-controlled region, it is likely they know more about Kaya victims than Borano victims. There may be many more victims throughout the country.
C1+T4	Numeric uncertainty, conference interval	Using these data, statisticians produced statistical estimates that suggest total Kaya victims range between 55,000 and 65,000, and that total Borano victims range between 2,000 and 38,000.

Vignette 3: Evidence Based-Sentencing

William dropped out of high school before graduating. He is now 22 years old living with his parents in a low-income housing project. He has developed a drinking problem, and his fed-up parents just told him he has until the end of the month to move out. He doesn't know how he will afford living on his own; he has been studying for the GED (a test that certifies high school-level academic skills), but he recently failed the GED exam. The next day, frustrated, he walks through a local mall looking for "help wanted" signs when he notices that a display case in a jewelry store has been left open. William waits until the salesclerk is distracted by a customer, reaches in to the case and sweeps a shelf full of diamond bracelets and necklaces into his backpack. He quickly walks away, but his crime was recorded by a security camera, and he is soon arrested. He pleads guilty to felony grand larceny (crime of high-value theft) in the amount of $150,000.

William's parents immediately kick him out of the house after his arrest. At the time of sentencing, he has been staying mostly with his older brother and with a high school friend, both of whom are known gang affiliates with drug distribution records. According to William's official Pre-sentence Report prepared by his probation officer, William has no prior felony convictions, but was convicted two years ago of misdemeanor drunk-and-disorderly conduct and underage-drinking charges.

[*C1+ Treatment here]

The statutory sentencing range for grand larceny in Michigan is 0 to 20 years' incarceration; actual sentences average 5 years, but they vary widely depending on the circumstances of the case.

If you were the sentencing judge, based on the information provided, how long an incarceration sentence would you give William?

____

(Type the number of years you think William should be sentenced to incarceration for his crime)

Table 10: Treatment language Vignette 3

Treatment	Name	Uncertainty message [*]
C0	No numbers	[No C1 paragraph above]
C1	Number, NO uncertainty	William’s Pre-sentence Report includes his score on the Risk of Reoffending Prediction Instrument (RRPI), which calculates his probability to commit another crime within the next three years. Scores range between 1 and 10: scores between 1-4 are considered “low risk,” 5-7 are considered “moderate risk,” and 8-10 are considered “high-risk.” William’s RRPI score is 8, placing him in the "high risk" category. The RRPI was developed by criminologists and uses available data about other people who have recommitted crimes after a first offense. This includes past offenders’ age, criminal record, education, marital status, employment status, stability of housing, substance abuse history, and their association with other people who have criminal records.
C1+T1	Simple Caveat	Note: The RRPI scores are calculated using existing and available information about offenders known to the authorities and well as known re-offenses. Thus, the data used is likely not to represent all offenders and their behaviors.
C1+T2	Rich Background	Note: The data used to calculate the RRPI scores come from crime records available to authorities on people who have committed multiple crimes. The data include people who have been arrested by law enforcement officers, charged with a crime at one time, and then at a later date, re-identified and formally charged with another crime. To be included in the data, police officers must have witnessed or been alerted to each of the crimes. They must have then filed accurate reports for the multiple crime committed by the same offender. The data must be properly entered into police databases. Further, multiple crimes must be properly linked to the same person. The data used to calculate the RRPI score are taken from the past 12 years, from offenders in the State of Michigan.

C1+T3	Expert Interception	Note: The data used to calculate the RRPI score come from available crime records on people who have committed multiple crimes over time. However, the available data may be biased, as they do not necessarily represent all offenders and their behaviors. We know, for example, that there are more data available about low-income offenders than high-income offenders who commit the same crimes. We also know that people with certain characteristics are arrested more often than other people for committing the same crimes. Given these biases in the available data, people from low-income demographics are more likely to be classified as high risk without necessarily being so.
C1+T4	Numeric uncertainty, conference interval	Note: Inaccuracies in the underlying data about past offenders, or minor differences in the defendant's age or housing stability, could make William's RRPI score be anywhere between 3 and 8.

References

Alston, P., & Knuckey, S. (2016). *The Transformation Of Human Rights Fact-Finding*. New York: Oxford University Press.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, *20*(3), 973–989.

Andreas, P., & Greenhill, K. M. (2010). *Sex, Drugs, and Body Counts : The Politics of Numbers in Global Crime and Conflict*. Ithaca, US: Cornell University Press.

Babbie, E. R. (2010). *The Practice of Social Research.* Belmont, Calif: Wadsworth Cengage.

Ball, P. (1996). *Who Did What to Whom? Planning and Implementing a Large Scale Human Rights Data Project*. American Association for the Advancement of Science.

Ball, P., Asher, J., Sulmont, D., & Manrique, D. (2003). *How Many Peruvians Have Died? An Estimate of the Total Number of Victims Killed or Disappeared in the Armed Internal Conflict between 1980 and 2000.* American Association for the Advancement of Science.

Ball, P., & Price, M. (2018). Annual Review of Statistics and Its Application Using Statistics to Assess Lethal Violence in Civil and Inter-State War.

Ball, P., & Price, M. (2019). Using Statistics to Assess Lethal Violence in Civil and Inter-State War. *Annual Review of Statistics and Its Application*, *6*(1), 63–84.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–

68.

Best, J. (2012). *Damned Lies and Statistics: Untangling Numbers from Media, Politicians, and Activists*. University of California Press.

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679.

Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review*, *82*(5), 977–1008.

Budescu, D. V, Por, H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climate Change*, *113*(2), 181–200.

Budescu, D. V, & Wallsten, T. S. (1987). Subjective estimation of precise and vague uncertainties. In G. Wright & P. A. yton (Eds.), *Judgmental Forecasting* (pp. 63–81). Wiley.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data. *Perspectives on Psychological Science*, *6*(1), 3–5.

Chanthika, P. (2006). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Charmaz, K. (2006). *Constructing Grounded Theory*. London ; Thousand Oaks, Calif: Sage Publications.

Christensen, L. T., & Cheney, G. (2015). Peering into Transparency: Challenging Ideals, Proxies, and Organizational Practices. *Communication Theory*, *25*(1), 70–90.

Cingranelli, D. L., & Richards, D. L. (2010). The Cingranelli and Richads (CIRI) Human Rights

Data Project. *Human Rights Quarterly*, *32*.

CNMH: CENTRO NACIONAL DE MEMORIA HISTÓRICA. (2013). *Basta Ya! Colombia: Memorias de Guerra y Dignidad*.

Cohen, D. K., & Green, A. H. (2012). Dueling incentives. *Journal of Peace Research*, *49*(3), 445–458.

Coomans, F., Grunfeld, F., & Kamminga, M. (2009). *Methods of Human Rights*. Antwerp, Oxford, Portland: Intersentia.

Correll, M., & Gleicher, M. (2014). Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2142–2151.

Creswell, J. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (3rd ed.). USA: Sage.

Creswell, J., & Clark, V. P. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). USA: Sage Publications.

Crismore, A., & Kopple, W. J. (1988). Reader's learning from prose: The effects of hedges. *Written Communication*, *5*, 184–202.

Dancy, G. (2016). Human rights pragmatism: Belief, inquiry, and action. *European Journal of International Relations*, *22*(3), 512–535. https://doi.org/10.1177/1354066115600038

Davenport, C., & Ball, P. (2002). Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977–1996. *Journal of Conflict Resolution*, *46*(3), 427–450.

Dieckmann, N. F., Mauro, R., & Slovic, P. (2010). The Effects of Presenting Imprecise Probabilities in Intelligence Forecasts. *Risk Analysis*, *30*(6), 987–1001.

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of Bias a Unified approach for Understanding Problems With risk assessment. *CRIMINAL JUSTICE AND BEHAVIOR*, *46*(2), 185–209.

Ellsberg, D. (1961). *RISK, AMBIGUITY, AND THE SAVAGE AXIOMS*.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes*, *45*(1), 1–18.

Fariss, C. J. (2014). Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *American Political Science Review*, *108*(2), 297–318.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fernandes, M., Walls, L., Munson, S., Hullman, J., & Kay, M. (2018). Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Human Factors in Computing, CHI '18*.

Fischhoff, B. (2012). Communicating Uncertainty: Fulfilling the Duty to Inform. *Issues in Science and Technology*, *28*, 63–70.

Frederick, S. (2005). Cognitive Reflection and Decision Making . *The Journal of Economic Perspectives*, *19*(4), 25–42.

Gächter, S., Orzen, H., Renner, E., & Starmer, C. (2009). Are Experimental Economists Prone to Framing Effects? A Natural Field Experiment. *Journal of Economic Behavior & Organization*, *70*(3), 443–446.

Gleditsch, K. S., Metternich, N. W., & Ruggeri, A. (2014). Data and progress in peace and conflict research. *Journal of Peace Research*, *51*(2), 301–314.

Goldstein, R. J. (1986). The Limitations of Using Quantitative Data in Studying Human Rights Abuses. *Human Rights Quarterly*, *8*, 607.

Greenhill, K. (2010). Counting the Costs: The Politics of Numbers in Armed Conflict. In P. Andreas & K. M. Greenhill (Eds.), *Sex, Drugs, and Body Counts : The Politics of Numbers in Global Crime and Conflict*. Ithaca, US: Cornell University Press.

Greis, M., Hullman, J., Correll, M., Kay, M., & Shaer, O. (2017). Designing for Uncertainty in HCI: When Does Uncertainty Help? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 593–600).

Guberek, T., & Hedstrom, M. (2017). On or off the record? Detecting patterns of silence about death in Guatemala's National Police Archive. *Archival Science*, *17*(1).

Guzmán, D., Guberek, T., & Price, M. (2012). *Unobserved Union Violence: Statistical Estimates of the Total Number of Trade Unionists Killed in Colombia,1999-2008*. The Benetech Human Rights Program.

Hall, K., & Stahl, D. (2008). *An Argument for Documenting Casualties: Violence Against Iraqui Civilians*. National Defense Research Institute.

Hannah-Moffat, K. (2013). Actuarial Sentencing: An "Unsettled" Proposition. *Justice Quarterly*, *30*(2), 270–296.

Heinzelman, J., & Meier, P. (2013). Crowdsourcing for Human Rights Monitoring: Challenges and Opporunities for Infomration Collection and Verification. In J. Lannon & E. Halpin (Eds.), *Human Rights and Information Communication Technologies: Trends and Consequences of Use*. USA: Information Science Reference, IGI Global.

Hullman, J. (2016). Why Evaluating Uncertainty Visualization is Error Prone. In *Association for Computing Machinery (ACM)* (pp. 143–151).

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, *25*(1), 903–913.

Human Rights Watch (2012). A Red Herring: Marijuana Arrestees Do Not Become Violent Felons.

Jasanoff, S. (2014). Serviceable Truths: Science for Action in Law and Policy. *Texas Law Review*, *93*.

Jensen, J. D. (2008). Scientific Uncertainty in News Coverage of Cancer Research: Effects of Hedging on Scientists' and Journalists' Credibility. *Human Communication Research*, *34*, 347–369.

Joslyn, S., & LeClerc, J. (2013). Decisions With Uncertainty: The Glass Half Full. *Current Directions in Psychological Science*, *22*(4), 308–315.

Joslyn, S., Nadav-Greenberg, L., & Nichols, R. M. (2009). Probability of precipitation: Assessment and enhance- ment of end-user understanding. *Bulletin of the American Meteorological Society*, *90*, 185–193.

Kadane, J. B., & Koehler, J. J. (2018). Certainty and Uncertainty in Reporting Fingerprint Evidence. *Daedalus*, *147*(4).

Kahan, D. M. (2011). Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law. *HARVARD LAW REVIEW*, *125*(19).

Kahan, D. M., Hoffman, D., Evans, D., Devins, N., Lucci, E., & Cheng, K. (2016). "Ideology" Or "Situation Sense"? An Experimental Investigation Of Motivated Reasoning And

Professional Judgment. *Univ. Pa. L. Rev*, *64*.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Keck, M. E., & Sikkink, K. C. N.-J. F. 529 . K. 1998. (1998). *Activists beyond borders: advocacy networks in international politics*. Ithaca, N.Y.: Cornell University Press.

King, G., Keohane, R., & Verba, S. (1994). Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton Univeristy Press.

Kruger, J., Ball, P., Price, M., & Green, A. H. (2013). It Doesn't Add Up: Methodological and Policy Implications of Conflicting Casualty Data. In T. Seybolt, J. D. Aronson, & B. Fischhoff (Eds.), *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*.

Kruger, J., & Lum, K. (2015). The Bodies of Evidence: An Exploration of Multiple Systems Estimation for Empirical Research with Violent Deaths. In *Midwest Political Science Association*.

Landman, T., & Godhes, A. (2013). A Matter of Convenience. In T. B. Seybolt, J. D. Aronson, & B. Fischhoff (Eds.), *Counting civilian casualties: an introduction to recording and estimating nonmilitary deaths in conflict*. Counting civilian casualties: an introduction to recording and estimating nonmilitary deaths in conflict: Oxford University Press.

Langford, M., & Fukuda-Parr, S. (2012). The Turn to Metrics. *Nordic Journal of Human Rights*, *30*.

Leiby, M. (2012). The Promise And Peril Of Primary Documents: Documenting Wartime Sexual Violence In El Salvador And Peru. In M. Bergsmo, A. B. Skre, & E. J. Wood (Eds.), *Understanding and Proving International Sex Crimes* (pp. 315–366).

List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what?

Some simple rules of thumb for optimal experimental design. *Experimental Economics*, *14*, 439–457.

Lum, K. (2017). Limitations of mitigating judicial bias with machine learning. *Nature*, *26*.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19.

Lupia, A. (2013). Communicating science in politicized environments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 14048–14054.

Lupia, A., & Elman, C. (2014). Openness in Political Science: Data Access and Research Transparency. *PS: Political Science & Politics*, *47*(01), 19–42.

Manski, C. (2018). The Lure of Incredible Certitude. *NBER Working Papers*.

March, J. G., & Simon, H. A. (1958). *Organizations*. Oxford, England: Wiley.

Marquart, F. (2017). Methodological Rigor in Quantitative Research. In *The International Encyclopedia of Communication Research Methods* (pp. 1–9). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (2013). The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength, and the weak evidence effect. . *Law and Human Behavior*, *37*(3), 197–207.

Martire, K., Kemp, R., Sayle, M., & Newell, B. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*(61–68).

McEntire, K., Leiby, M., & Krain, M. (2015a). How to Ask People for Change: Examining Peoples' Willingness to Donate to Human Rights Campaigns. In J. R. Pruce (Ed.), *The Social Practice of Human Rights*. New York: Palgrave Macmillan.

McEntire, K., Leiby, M., & Krain, M. (2015b). Human Rights Organizations as Agents of

Change: An Experimental Examination of Framing and Micromobilization. *American Political Science Review*, *109*(3), 407–426.

Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human rights, Gender Violence, and Sex Trafficking*. Chicago : The University of Chicago Press, 2016.

Merry, S. E., Davis, K. E., & Kingsbury, B. (2015). The Quiet Power of Indicators. New York: Cambridge Univeristy Press.

Moyn, S. (2010). *The Last Utopia: Human Rights in History*. Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability.

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown.

Oreskes, N., & Conway, E. M. (2010). *Merchants of Doubt: How a handful of scientists obsured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury Press.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Peirce, C. (1966). *Selected Writings (Values in a Universe of Chance)*. New York: Dover Publications, Inc.

Porter, T. (1995). *Trust in Numbers : The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press.

Powers, M. (2016). A new era of human rights news? Contrasting two paradigms of human rights news-making.

Price, M., & Ball, P. (2015a). Selection bias and the statistical patterns of mortality in conflict 1.

Statistical Journal of the IAOS, *31*, 263–272.

Price, M., & Ball, P. (2015b). The Limits of Observation for Understanding Mass Violence. *Canadian Journal of Law and Society*.

Retzbach, A., & Maier, M. (2015). Communicating Scientific Uncertainty: Media Effects on Public Engagement With Science. *Communication Research*, *42*(3), 429–456.

Reydams, L. (2016). NGO Justice: African Rights as Pseudo-Prosecutor of the Rwandan Genocide. *Human Rights Quarterly*, *38*(3), 547–588. Retrieved from https://muse.jhu.edu/article/627627

Root, B. (2016). Numbers are Only Human: Lessons for Human Rights Practitioners from the Quantitative Literacy Movement. *The Transformation of Human Rights Fact-Finding*.

Rosga, A., & Satterthwaite, M. L. (2009). The Trust In Indicators: Measuring Human Rights. *Berkeley Journal of International Law*, *27*(2), 253.

Roth, F., Guberek, T., & Green, A. H. (2011). *Using Quantitative Data to Assess Conflict-Related Sexual Violence in Colombia: Challenges and Opportunities*. Palo Alto and Bogotá: Benetech Human Rights Program and Corporación Punto de Vista.

Satterthwaite, M. L., & Simeone, J. C. (2016). An Emerging Fact-Finding Discipline? A Conceptual Roadmap for Social Science Methods in Human Rights Advocacy. In *The Transformation of Human Rights Fact-Finding*. Oxford : Oxford University Press, 2016.

Seybolt, T. B., Aronson, J. D., & Fischhoff, B. C. N.-U. 21. . . C. 2013. (2013). *Counting civilian casualties: an introduction to recording and estimating nonmilitary deaths in conflict*. *Studies in strategic peacebuilding*. Counting civilian casualties: an introduction to recording and estimating nonmilitary deaths in conflict: Oxford University Press.

Sharp, D. N. (2018). Pragmatism and Multidimensionality in Human Rights Advocacy. *Human*

Rights Quarterly, *40*(3), 499–520. https://doi.org/10.1353/hrq.2018.0029

Simon, H. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, *63*(2), 129–138.

Slovic, P., Zionts, D., Woods, A. K., Goodman, R., & Jinks, D. (2013). Psychic Numbing and Mass Atrocity. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy*. Princeton Univeristy Press.

Small, D. A., & Verrochi, N. M. (2009). The Face of Need: Facial Emotion Expression on Charity Advertisements. *Journal of Marketing Research*, *46*(6), 777–787.

Starr, S. B. (2014). Evidence-Based Sentencing and the Scientific Rationalization of Discrimination. *Stanford Law Review*, *66*(4), 803–872.

Tate, W. (2007). *Counting the Dead: The Culture and Politics of Human Rights Activism in Colombia*. Berkeley and Los Angeles: University of California Press.

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In *8th International AAAI Conference on Weblogs and Social Media*.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Valdez, P., & Mehrabian, A. (1994). Effects of Color on Emotions. *Ournal of Experimental Psychology: General*, *123*(4), 394–409.

Vivalt, E., & Coville, A. (2018). *How Do Policymakers Update?*

Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *The Knowledge Engineering Review*, *10*(1), 43–62.

Weidmann, N. B. (2016). A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science*, *60*(1), 206–218.

Williams, H. E., Bowman, S. W., & Jung, J. T. (2016). The Limitations of Government Databases for Analyzing Fatal Officer-Involved Shootings in the United States. *Criminal Justice Policy Review*.