EARLY WARNING PROJECT

TAKING STOCK OF EARLY WARNING FOR ATROCITY PREVENTION

2021 Sudikoff Interdisciplinary Seminar Series on Genocide Prevention

Rapporteur's Reports | June 2021

UNITED STATES HOLOCAUST MEMORIAL MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

Statistical Methods for Early Warning and Risk Assessment

Mollie Zapata
Simon-Skjodt Center for the Prevention of Genocide
US Holocaust Memorial Museum
17 June 2021

On June 17, 2021, the Simon-Skjodt Center for the Prevention of Genocide at the U.S. Holocaust Memorial Museum brought together scholars, practitioners, and civil society representatives to discuss statistical methods for early warning and risk assessment. This seminar was the first in a series designed to stimulate discussion about the current state of early warning for atrocity prevention. Motivations for this seminar included (1) sharing lessons from and with the Early Warning Project (EWP) to improve the field of early warning, (2) improving early warning and response mechanisms within the U.S. government, and (3) generating new ideas for future research to improve the field of early warning. Other seminar series topics included qualitative early warning assessments, and communicating risk. This rapporteur's report summarizes major observations raised during the discussion under the Chatham House rule of non-attribution and does not necessarily represent the views of all participants.

INTRODUCTION

Statistical forecasting techniques are central to contemporary early warning. Governments, civil society organizations, and independent researchers have built a variety of platforms to address the critical need for reliable, consistent, public early warning for catastrophic events, including mass atrocities. Alongside qualitative approaches, these projects use varying statistical methods and data inputs. The U.S. Holocaust Memorial Museum and Dartmouth College founded the Early Warning Project in 2014. As the field of early warning develops, it is critical that researchers regularly review the state of the science to ensure high-quality, impactful analysis. Consistent with its goal of advancing the science of early warning, the Early Warning Project is currently reviewing the possibilities to revise its methods and data in 2022. Discussions amongst researchers, especially those that include key audiences (i.e. policymakers and civil society organizations), are critical to f advancing the field and identifying avenues for future research.

In this 90-minute virtual convening, participants reflected on challenges they have faced in quantitative risk forecasting, covering current and emerging developments in data sources and methods, as well as use and communications challenges.

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

THE CURRENT STATE OF DATA AND METHODS

Participants noted that the atrocity forecasting field should continue to refine its methods by (1) improving applications of historical data, and (2) continuing to refine the statistical models that it uses to estimate the risk of future mass atrocity onsets.

Data: retrospective coding

In the discussion of data sources, participants debated how projects should deal with sources that make retrospective changes to their coding. Many crossnational datasets, for example, will revise their estimate of variables such as "level of democracy" to reflect changes that became apparent only after initial codings were made. Participants discussed whether retrospective coding changes result in improvements to the training set, thereby improving forecasting outcomes, or whether these changes lead to data biases (e.g., if coders rate countries that experienced mass atrocities as having been less democratic than they would have absent knowledge of the subsequent mass atrocities) or over-fitting of prediction models. One scholar noted that this affects a relatively small universe of country-year coding, and that atrocity forecasting researchers should expect any changes to fall within reasonable confidence bounds. However, some initial research suggests that retrospective coding change rates were higher in countries that had experienced mass violence than those that had not.

Data: gaps and temporal challenges

There are two enduring data challenges that all projects working on forecasting face: (1) the scope of data available, and (2) the time lag between data collection, processing, and the production of risk assessments.

Most of the forecasting efforts discussed rely on independent sources for measures of predictor variables, and are therefore beholden to the regional and temporal constraints and data collection approaches of other organizations. While many new crossnational datasets have been developed in recent years, relatively few cover the global scope and time period that forecasting projects focusing on rare events require. Further, researcher biases and other obstacles to data collection and creation may affect the measurement of important predictor variables. Participants noted that though most high-risk countries are in sub-Saharan Africa, research about countries in this region often faces political and technological constraints. This may lead to unreliable or inaccurate data.

Regarding timing, most datasets are compiled annually, posing a challenge to those seeking to influence the fast tempo of policy decision-making. Some efforts are underway to build datasets using different sources and methods that may update predictors and violence outcomes on a more frequent basis. In particular, participants noted an effort to collect publicly available press releases and blog posts from human rights NGOs. This project uses unstructured language and sentiment analysis to train a machine-learning model that identifies perpetrators and victims of specific human rights violations in near-real-time.

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

One contributor asked about the extent to which civil society actors in high-risk countries are involved in developing early warning tools and methods. Without sufficient inclusion of these actors, all aspects of mass atrocity forecasting—the data collection, model development, and communication—may be adversely impacted.

Regarding data sources, participants identified a number of gaps in the field and some new efforts underway. These include data to track (1) gender, (2) hate speech (especially in languages other than English), (3) events that may trigger mass atrocities, (4) ecological developments resulting from climate change and related phenomena, (5) sub-national patterns of violence and human rights violations, (6) horizontal inequalities, and (7) regional and ethnic affiliations of political leaders. Participants also noted projects that measure human rights violations and protections, and a global news monitoring system designed to monitor ingroup/outgroup conflict discourse. Participants noted potential applications of advanced modeling techniques to address data gaps, including text mining, machine learning using social media data, and hybrid "human / machine" collection of atrocity information.

Modeling: types of models

Most of the projects discussed use some version of a logistic regression model that estimates the probability of the binary onset or non-onset of an event (i.e., mass killing or genocide/politicide) in a given time period. Most participants said they had not found more complex models to be significantly more accurate on these types of forecasting tasks.

One participant noted that, although the discussion had focused on input variables, more thought should be put towards what outcome the models are forecasting. Returning to the discussion of data gaps, there is a need for more complete data on and forecasting efforts towards atrocity crimes that occur outside battlefield contexts, such as imprisonment and forced sterilization.

Modeling: performance metrics and dealing with uncertainty

Testing and reporting on uncertainty and various model performance measures is a challenge that many projects approach differently. One of the challenges in working with rare events is that most standard performance measures are not helpful in this case (i.e. the "area under the receiver-operating characteristic curve" typically does not represent the performance of a rare event forecasting model well).

Some participants deal with this by reporting the difference between average estimated risk for countries that experienced an event and those that did not; the greater the difference, the better the model is performing. Models for rare events tend to have low precision, meaning there are a large number of false positives (forecasted atrocity episodes that do not occur) for every true positive. One participant suggested using simulations to deal with uncertainty. Participants noted that researchers working on forecasting should explore ways to describe uncertainty to policymakers and the public without undermining the forecast findings or downplaying the need for preventive action. Some projects choose not to present confidence intervals for risk scores

for fear that non-researchers may misinterpret uncertainty and devalue the analysis as a result.

USE AND COMMUNICATIONS CHALLENGES

Taking a step back from the technical details, participants recommended more scrutiny of whether projects are accomplishing their early warning and prevention objectives. Participants suggested additional research into whether early warning "failures" should be attributed to model performance or bureaucratic and communications challenges. They also asked whether, or to what extent, various forecasting methods may pick up successful policy efforts to prevent mass atrocities that resulted from correct forecasts.

Participants with policy experience indicated that the direct policy implications of forecasts are of utmost importance. They also suggested that the connection between early warning assessment and preventive diplomatic activities needs to be further explored. On this matter, some participants asserted that the purpose of these analytic tools is to prod people in government to direct attention on countries at risk. An additional policy challenge raised was that when the key audience is accustomed to relying on qualitative information, analysts must explore ways to ensure understanding and trust of quantitative methods.

Additionally, participants noted other issues of how different audiences interpret quantitative forecasts. The models discussed are producing forecasts rather than causal explanations, raising a question of how to communicate forecasting coefficients to the public while underscoring that these are not causal variables. Beyond the technical aspects of evaluating accuracy and precision, participants noted that perhaps the more important question is how many countries policymakers have bandwidth to address, and how many countries should be included on "risk lists" as a result of these bandwidth issues. Participants underscored that these lists should have the highest recall possible, subject to limitation on attention rather than producing risk lists that will be dismissed because they are too long. In other words, as many countries should be included on risk lists as policymakers can feasibly understand and address.

QUESTIONS FOR FURTHER RESEARCH

- What types of forecasting information are the most useful for policymakers? For country-level forecasts, how many countries should forecasting reports include? At what level of precision should researchers present forecasting findings?
- What new crossnational data inputs are useful predictors of mass atrocities?
- What do advances in latent variable models contribute to the methods and outputs of the risk forecasting field?

Qualitative Early Warning Assessments
Rapporteur's Report

Tahlia Mullen
Simon-Skjodt Center for the Prevention of Genocide
US Holocaust Memorial Museum
22 June 2021

On June 22, 2021, the Simon-Skjodt Center for the Prevention of Genocide at the U.S. Holocaust Memorial Museum brought together scholars, practitioners, and civil society representatives to discuss qualitative early warning assessments, with a focus on the State Department/USAID Atrocity Assessment Framework. This seminar was the second in a series designed to stimulate discussion about the current state of early warning for atrocity prevention. Motivations for this seminar included (1) sharing lessons from and with the Early Warning Project (EWP) to improve the field of early warning, (2) improving early warning and response mechanisms within the U.S. government, and (3) generating new ideas for future research to improve the field of early warning. Other seminar series topics included statistical methods for early warning and risk assessment, and communicating risk. This rapporteur's report summarizes major observations raised during the discussion under the Chatham House rule of non-attribution and does not necessarily represent the views of all participants.

INTRODUCTION

As global actors endeavor to prevent new mass atrocities, qualitative risk assessments provide one promising opportunity to identify and communicate early warning signs. The Department of State/USAID Atrocity Assessment Framework (AAF) supplies practitioners with guidance to conduct such assessments. Created in 2015, the AAF has been successfully implemented in risk assessments conducted by both government and non-governmental organizations.[1] Although researchers and practitioners alike regard the Framework as an effective tool for qualitative analysis of mass atrocities, the Framework also faces challenges, suggesting there is room for future improvement.

During the seminar, participants reflected on their experiences designing, executing, and communicating results of early warning analyses conducted using the AAF. The conversation highlighted features perceived to be most useful to policymakers, such as the sections on windows of opportunity and recommendations for action against mass atrocities. Participants also discussed barriers to effectiveness, such as difficulties achieving buy-in from both local and international actors, and making assessments and their resulting recommendations accessible

---

[1] The Simon-Skjodt Center provided partial support for the Framework's development via a visiting fellowship to Annie Bird, then a policy advisor with the Department of State's Bureau of Conflict and Stabilization Operations.

and attractive to policymakers inside the U.S. foreign policy bureaucracy. In the course of discussion, participants noted that certain opportunities for improvement come with trade-offs that must be weighed ahead of implementation. Dialogue between participants supplied both possible solutions to issues raised over the course of discussion and highlighted priority areas for future improvement.

UTILITY FOR RESEARCHERS AND SCHOLARS

Participants largely agree that the AAF provides researchers and scholars with a useful analytic tool for conducting comprehensive analyses of risk. One researcher praised its emphasis on potential perpetrators, targets, and triggers as a basis for understanding how mass atrocities occur. Another noted its attention to risk dynamics, plausible scenarios, and "key developments" to monitor as a useful guide for officials implementing the AAF in practice. For example, one participant noted that when violence broke out in Burundi in 2015, the risk assessment's emphasis on these factors to watch enabled more rapid U.S. government allocation and distribution of resources, such as additional programmatic assistance to civil society groups.

A second strength of the AAF is its discussion of "windows of opportunity," which are defined as changes in situations that increase incentives for perpetrators to turn away from atrocities or for local or international groups to prevent or respond to atrocities. A participant noted that referencing windows of opportunity enables analysts to make targeted recommendations for preventive action.

POLICY AND PUBLIC PERCEPTION CHALLENGES

Several participants noted that while the AAF creates a clear basis for risk analysis, it does not provide detailed guidance for the development or focus on policy recommendations. That weakness matters because policymakers are most interested in exploring solutions to risks of mass atrocities. One researcher noted that the AAF does not indicate whether it is a tool for policy planning. Although it is meant to complement Conflict Assessment Frameworks that U.S. government agencies use to analyze the general risk of violent conflict in particular countries, it is unclear how to use the two frameworks together.

A second challenge in implementing the AAF is that explicit discussions of "mass atrocities" and potential perpetrators can elicit negative reactions from the governments of countries under focus. Often, these governments may contribute to mass atrocity risks or bear responsibility for large-scale human rights abuses that practitioners seek to prevent. When a risk assessment identifies the ruling party or security forces as a potential perpetrator, the government and its allies in civil society may work to delegitimize the report by publicly denouncing it or accusing international actors of interfering in the country's internal affairs. One participant who observed the AAF struggling to achieve mainstream acceptance in Jakarta suggested improving local buy-in by changing some of the terminology to reflect more politically palatable framing for the local context.

A final challenge of the AAF is making reports accessible and attractive to policymakers inside the U.S. foreign policy bureaucracy. No government wants to be told that what they are doing is not working. A senior policy advisor noted the presence of other high-priority crises and lack of resources are barriers to demonstrating the value of preventive action to policymakers.

OPPORTUNITIES FOR IMPROVEMENT

Much of the discussion of how best to improve the AAF centered on "bridging the gap" between risk analysis and policy implementation. Discussants agreed that simply understanding risk is insufficient for preventing mass atrocities. One researcher recommended that instead of focusing on risk assessment, the AAF should focus on assessing the window of opportunity for prevention. The revised purpose of the AAF could be to help policymakers understand atrocity dynamics while also considering potential responses and their chances for success. Another participant recommended that analysts remain in continuous communication with the NSC and State Department throughout the analysis. This approach helped make countries like Burundi, which are traditionally sidelined in Washington, a much higher priority for policymakers.

Other opportunities for improvement centered on uptake of the results of the AAF among practitioners outside Washington. One analyst recommended communicating sensitive results of risk assessments through private channels to key policymakers and translating the AAF and assessment into the local language, potentially forgoing the use of controversial terms (including "genocide" and "mass atrocities"). Another recommendation was to acknowledge existing preventive efforts in the region and frame the risk assessment analysis in ways that are useful to these efforts. A third opportunity for improvement is reforming the budgeting process. By aligning this process more closely with analysis, atrocity prevention resources may be allocated more efficiently and have more direct impact in countries under focus.

Participants brought up numerous other opportunities for improving risk analysis and policy implementation, including:
- Greater focus on gender dynamics and considerations
- More attention to monitoring to track the trajectories of key countries as mass atrocity risks change
- Greater focus on the role of law enforcement and judicial systems in encouraging and preventing potential mass atrocities
- Attention to how climate-related issues exacerbate violence
- More analysis of structural and political factors contributing to violence, including a more rigorous focus on third party disablers
- Ensuring a diverse team of analysts with deep local expertise are charged with conducting assessments

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

MANAGING TRADE-OFFS

While considering present challenges and opportunities for future improvement, participants identified what trade-offs must be weighed ahead of making permanent changes to the AAF. One dilemma is whether or not to make risk assessments mandatory. A participant argued that mandated usage would help attract greater attention from policymakers and thus increase resource allocation; however, efforts to enforce such a mandate could distract from achieving success on the ground. A second dilemma is whether to keep the results of risk assessments public or private. Which approach is more advantageous depends on the country, the researchers, and the findings of the analysis. A third dilemma is how to ensure bureaucratic attractiveness while maintaining the integrity of the analysis. One participant noted that researchers might feel pressure to revise their analysis to improve its appeal to NSC and State Department officials. A final dilemma is whether adding additional factors of analysis leads to a deeper understanding of the issue or rather presents a distraction. A participant with experience using the AAF noted the fine line between directing the focus of analysis versus ignoring key information.

QUESTIONS FOR FURTHER RESEARCH

- How can more diverse sources of data be integrated into analysis? How do we determine what information is and isn't relevant for a given case?
- How can the Atrocity Assessment Framework be used in conjunction with traditional conflict assessments?
- How might the bureaucratic and career risks that come with conducting and releasing results from qualitative risk assessments be mitigated?
- What can be done to ensure that indicators used in risk assessments by analysts are framed in terms that are meaningful to policymakers?

Communicating Risk
Rapporteur's Report

Tahlia Mullen
Simon-Skjodt Center for the Prevention of Genocide
US Holocaust Memorial Museum
24 June 2021

On June 24, 2021, the Simon-Skjodt Center for the Prevention of Genocide at the U.S. Holocaust Memorial Museum brought together scholars, practitioners, and civil society representatives to discuss communicating risk for early warning and mass atrocities assessments. This seminar was the third in a series designed to stimulate discussion about the current state of early warning for atrocity prevention. Motivations for this seminar included (1) sharing lessons from and with the Early Warning Project (EWP) to improve the field of early warning, (2) improving early warning and response mechanisms within the U.S. government, and (3) generating new ideas for future research to improve the field of early warning. Other seminar series topics included statistical methods for early warning and risk assessment, and qualitative early warning assessments. This rapporteur's report summarizes major observations raised during the discussion under the Chatham House rule of non-attribution and does not necessarily represent the views of all participants.

INTRODUCTION

Risk communication is a critical component of early warning systems and mass atrocity prevention. Efforts towards improving the quality of risk assessments matter only to the extent that the assessments enhance policymakers' understanding of risk, conception of opportunities for prevention, and motivation towards action. Because policymakers rarely approach risk assessments with attention towards research challenges, they rely on researchers to conduct analyses and effectively communicate findings about mass atrocity risks.

Seminar participants discussed the status of risk communication at present, identifying barriers that prevent clear and coherent analysis from reaching policymakers or having its intended impact. A recurrent theme of discussion was how a combination of political incentives, status-quo biases, and groupthink can lead policymakers to question or reinforce the reliability of risk assessments. Beliefs and biases may prevent officials from "speaking truth to power" and may instead lead them to overlook or ignore reasonable estimates of mass atrocity risks. Participants argued that improving risk communication requires overcoming such biases, both by improving the quality of warnings and by designing messages that maximize policymakers' willingness to receive them.

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

HOW POLITICAL OFFICIALS UNDERSTAND RISK

As one scholar summarized, the failure to act on warnings is most often due to the interaction between a lack of quality warning and a lack of political will to heed warnings when they do occur. These two pieces are greatly intertwined: for example, warnings that ignore the role of cognitive biases in policy decision-making may not elicit an adequate response from policymakers. A second scholar observed that understanding risk is more a feeling than a fact. Consequently, how much policymakers understand risk depends on their level of personal, bureaucratic, or political investment in the issue in question. Those favoring a particular action are more likely to overlook its risks, whereas those indifferent or opposed to an action are more likely to exaggerate its risks. This participant also observed that when officials perceive actions to prevent mass atrocities as adversely impacting their own national security, national security will take priority.

Because the expertise of policymakers lies in crafting and implementing policy strategies and not in risk assessment, they are prone to misinterpret the results of both quantitative and qualitative analyses of mass atrocity risk. For instance, when told a particular outcome is a worse case scenario, officials tend to equate "worst case" with "unlikely," regardless of how probable the outcome is. Officials are also prone to conflate risk with uncertainty, two distinct concepts which deserve to be treated as such. In short, the language that imparts one meaning to scholars may impart a very different meaning to the policymakers who review and act on warnings.

IMPROVING WARNING QUALITY

Given the role risk assessments play in determining a policymaker's understanding of risk, improving the quality of warning is central to overcoming the so-called "warning-response gap." This gap is not just about the space between analysis and policy, but also between the distinct logics of warning and of prevention. Participants offered numerous strategies for improvement including:

- Providing greater insights into triggering events, particularly determining whether or not a particular event will likely lead to violence
- Focusing on risk potentials and the benefits of early action to help encourage prevention work
- Adjusting risk assessment methods to highlight tools to spot early warning signs of genocide
- Bridging miscommunications by creating more spaces and structures to connect individuals in different roles
- Developing a common language to ensure that warnings are not lost in translation

Improving policymaker's understanding of risk also requires improving credibility. One scholar recommended using training to spot disinformation, red teams, and

patience to help avoid reporting on false positives and ensure sufficient resources are given to legitimate threats. Another participant recommended providing officials with a track record indicating the past accuracy of a given source. Finally, spaces and structures could also be created to connect researchers, thereby improving the consistency of messages reaching policymakers and avoiding contradictions that undermine credibility.

IMPROVING MESSAGING

Overcoming the "warning-response gap" requires not only improving the quality of warnings but also the messages that analysts impart to policymakers. Participants argued that messages are more likely to engender response when they emphasize:

- Agency: the power to predict risk and prevent mass atrocities
- Urgency: the limited time frame during which early action is effective
- Obligation: the responsibility to prevent atrocities and avoid repeating past failures

Warnings are also more effective when they are tailored according to their specific objective and intended targets. In the U.S. government context, for example, analysts should be sensitive to the finding that language and concepts about mass atrocities resonate differently depending on the audience's party affiliation. Generally speaking, messages are most effective when they avoid jargon and use language accessible to individuals both within and outside the academic community.

One scholar argued that among the most important considerations of effective risk communication is deciding who is the correct spokesperson for a message. Researchers and academics are not necessarily the right people to communicate risk to policymakers. The task may require an intermediary whose central purpose is to translate analyses into a form comprehensible to policymakers. Additionally, warnings may have the greatest impact when they come from local voices. The closer an individual is to a problem, the more credible their message can be.

How messages are delivered is also worth considering. Analysts are not necessarily well connected with policymakers and may struggle to convey warning signals to those with the agency to act. A participant recommended that in such cases, circulating messages publicly through the media may help to overcome the bureaucratic barriers otherwise preventing the open flow of information.

QUESTIONS FOR FURTHER RESEARCH

- How can analysts make information about risk more precise and persuasive despite uncertainty about potential future outcomes?
- What role, if any, should digital and social media play in communicating risks of mass atrocities now and in the future?

- What strategies can help analysts balance their efforts to make early warning persuasive while remaining faithful to their analysis?
- How can warnings communicated to policymakers be useful to civilians at risk and how can we ensure the concerns of civilians at risk are effectively communicated to policymakers?