Patterns of Conflict and Cooperation in Liberia (Part 2): Prospects for Conflict Forecasting and Early Warning

Robert Blair
Yale University

Christopher Blattman, PhD, MPA/ID
Yale University & IPA

Alexandra Hartman
Yale University

Acknowledgements: The research was supported via grants from the United Nations Peacebuilding Fund in Liberia (via UNHCR), Humanity United, Yale University, the MacMillan Center, and the World Bank's Italian Children and Youth Trust Fund. We especially thank Jonathan Andrews, Mamadou Dian Balde, James Ballah, Wilfred Grey-Johnson, Jason Hepps, Mattias Lundberg, Michael Kleinman, Tomoko Semmyo, Chiara Tufarelli and the Norwegian Refugee Council for their assistance. In addition, we received valued comments on the research design and analysis from numerous scholars and practitioners, including participants in the Yale ISPS and OCV seminars. Finally, Bryan Plummer, Mathilde Emeriau, Tricia Gonwa, Brittany Hill, Rebecca Littman, Benjamin Morse, Johnny Ndebe, Gwendolyn Taylor, Prince Williams, and John Zayzay provided superb research assistance. Cover photo: Liberian women on their way to market in Voinjama, Lofa County.

Executive Summary

Conflict early warning remains an important but elusive goal in Liberia. If outbreaks of violence could be predicted before they occur, early responders could focus their energies and scarce resources on the highest-risk communities.

Is such a goal realistic? Early warning requires a simple system for generating reliable predictions—a system that is not only accurate, but is also consistent over time and across counties and communities. This is a difficult, maybe impossible, task.

In this report we describe results from a two-year study that suggest prediction may be more promising than we initially expected. We use fine-grained quantitative data from a survey of 247 rural Liberian towns and villages to assess whether statistical analysis can be used to predict conflict over time. To our surprise, we find that models built on fewer than 10 risk factors measured in 2008 accurately predict up to 75% of all conflicts two years later. We began this exercise skeptical, and these accuracy rates are far higher than we anticipated.

We also find that these models are flexible, and can be trained not only to maximize accuracy, but to minimize "false negatives" as well. The costs of conflict are enormous, and we believe it is preferable to predict that conflict will occur even when it doesn't than to predict it will *not* occur even when it does. Our models accurately predict 40 to 70 percent of all incidents of conflict ("true positives"), with three to five false alarms (or "false positives") for each correctly predicted incident.

Equally important, our models are able to distill a small subset of robust predictors from a long and unwieldy list of potential risk factors. Out of 60 potential risk factors, each model identifies 5 to 10 correlates that are especially powerful in predicting the onset of conflict, and that government and NGOs might use to target high-risk communities. These risk factors are not always consistent across models, but they tend to cluster around four common themes:

1. Population size;
2. Ethnic and religious diversity;
3. Presence of ex-combatants and exposure to wartime violence; and
4. Intra-communal wealth and assets

In our discussion we offer an interpretation of each of these four categories of risk based on casual observations and systematic, in-depth qualitative research in these 247 communities.

Finally, we conclude with recommendations for advancing the goal of conflict early warning in Liberia. First, we urge the early warning/early response (EWER) community to focus less exclusively on analysis of past conflicts and more on anticipation of future ones. Second, we encourage researchers and NGOs to standardize their indicators for conflict and risk. Third, to test whether these (or alternative) models are consistent outside our sample, we suggest three possible next steps for data-driven EWER:

1. Collect new data on these 247 communities to test the generalizability of the model to 2011 and 2012.
2. Test the consistency of the model to a nationally representative dataset, such as that collected by the Berkeley Human Rights Center.
3. Pilot a simple system for high-frequency data collection and forecasting in high-risk areas.

Contents

1. Overview: The search for early warning

As Liberia strengthens its security sector and rule of law, policymakers face hard questions about how best to allocate scarce resources. Where should police patrol? Which communities need the most assistance with conflict resolution and mediation? Where should new Peacebuilding Fund (PBF) monies be spent?

Meanwhile, the everyday business of peacekeeping and law enforcement requires a coherent, actionable system for responding to conflicts and preventing their escalation. Where to target and how to respond?

These questions are more easily answered if we can predict where violence is most likely to occur. Peacekeepers, government and civil society in Liberia have expressed a common interest in developing an early warning and early response (EWER) system to avoid the escalation of intra-communal conflicts into regional or national ones. Such a system not only would identify small-scale disputes as they emerge, but also would track risk factors and leading indicators in order to prevent escalation into larger-scale violence.

The first requirement—identifying conflicts as they emerge—is largely a technological and logistical challenge, with a host of emerging solutions. The second requirement—forecasting—raises a more fundamental question: is such a thing even possible? Even with the right data, is it possible to anticipate outbreaks of large-scale conflict—riots, strikes and protests, violent confrontations between ethnic or religious groups—*before* they occur?

In truth, no one knows. Forecasting the future is difficult, maybe impossible, and failed attempts at prediction may yield misguided policy conclusions. But forecasting is not necessarily impossible, and if it can be done, the returns for EWER are potentially enormous. This report constitutes a first step towards statistical forecasting and early warning in Liberia, harnessing data from a uniquely fine-grained survey on conflict and its correlates. The results, we discover, are promising.

2. Forecasting the future: What it entails

Forecasting can take many forms. One form—perhaps the most common—is based in theory, intuition or plain common sense. In the absence of empirical evidence, we might have strong theoretical priors about the events and circumstances most likely to cause conflict: a stolen election, for instance, or a call to arms by a popular political leader. We can use these priors to predict where we believe conflict is most likely to occur.

An alternative form is based on data. This data-driven approach is not necessarily devoid of theory, but it tends to be agnostic about what might or might not cause conflict, opting to let the data speak for themselves. It is this second approach that we attempt, informed by our theoretical priors but not bound to them.

Our goal is to design a statistical model that uses risk factors and leading indicators to predict where episodes of violence are most likely to erupt. The ideal model would meet at least four criteria:

- **Simplicity.** The list of potential risk factors must be manageably small. Complex models are likely to prove unwieldy for policymakers, and may yield less accurate predictions than their simpler counterparts.
- **Measurability.** Potential risk factors must be easily observable and measurable at relatively low cost. Models based on unmeasurable risk factors will not be tractable in practice.
- **Consistency.** The model must generalize over time and space. The less generalizable the model, the less useful it will be for real-world EWER.
- **Accuracy.** The model must reliably predict conflict, maximizing "true positives" and minimizing "false positives" and "false negatives." Inaccurate models may be worse than no model at all.

This is a tall order. The universe of potential risk factors is almost prohibitively large, and there exists almost no evidence to distinguish reliable from unreliable forecasters. Some of the risk factors that matter most for predicting violence—fear, anger, cultural and political norms—may also be the hardest to measure. Data-driven forecasting is paradoxically "backwards-looking," using past conflicts and their correlates to predict future ones—a fact that hinders consistency and accuracy. The world is constantly changing, and what predicted violence last year may not predict it in the future. An election, war in a neighboring country, a change in international law: all of these can fundamentally alter the dynamics of conflict and frustrate attempts at prediction. And these macro-level trends do not even take into consideration the hundreds of micro-level political, demographic, and institutional changes going on in a place like Liberia as it emerges from years of dictatorship and civil war.

These difficulties raise two more immediate and fundamental questions that must be answered before statistical EWER can proceed. From the universe of potential risk factors for conflict, can we identify and measure a subset of the most robust predictors? Can we use those predictors to build a model that "retro-casts" violence—in the past, in a single time period, and in a limited number of communities—more accurately than chance alone? If the answer to either is no, then the data-driven approach is almost certainly a dead end.

3. New and unique conflict data

A. Data

Over the past year, a research team from Yale University and Innovations for Poverty Action has attempted to answer these questions using fine-grained survey data from almost 1,000 local leaders and over 9,000 community members in 247 towns and villages in Lofa, Nimba, and Grand Gedeh counties.

The data were collected as part of our evaluation of one of Liberia's largest peacebuilding programs—the PBF-funded Community Empowerment Program.[1] Data collection occurred in two rounds:

- Round 1: In March and April of 2009, IPA interviewed a random sample of 20 adults in each of the 247 communities. The team also surveyed four leaders in each community—typically a town chief, a female leader, a youth leader and a minority leader—for a total of 5,632 respondents.
- Round 2: In November and December of 2010, IPA returned to each of the 247 communities and conducted a second round of surveying with a new random sample of 20 adults. The team also attempted to interview the same 4 leaders from Round 1,[2] for a total of 5,741 respondents.

Using these two rounds of surveying, we construct a panel dataset that draws on the same pool of leaders over time and a repeated cross-section of randomly-selected community members, representative of each of these 247 towns and villages.

In an earlier report we leveraged these data to examine patterns of conflict and cooperation in rural Liberia.[3] Some key findings include:

- Incidents of interpersonal and collective violence—assaults, mob justice, ritual killings—are destabilizing, but are relatively rare and appear to be decreasing in frequency over time.
- Disputes over land use, tenure and inheritance are pervasive. Many of these disputes never escalate into violence, but a surprisingly high proportion of them do.
- Intra-communal cohesion and trust—both important indicators of reconciliation—seem to be deteriorating. Inter-tribal biases and stereotypes remain endemic, perpetuating complaints of inequity among Liberia's ethnic minorities.

B. Our measure of conflict

In our two previous reports we examined a variety of different indicators for conflict, from petty interpersonal disputes to crime to riots and protests. In this report we shift focus towards a single indicator that we believe captures the most serious threats to stability in rural Liberia. For purposes of this report, we define "conflict" as any one of the following incidents:

---

[1] The Community Empowerment Program (CEP) involved a series of small, intensive, eight-day workshops conducted with roughly 10% of all adults in each community, typically over several weeks or months. The aim of the program was to: (i) educate people on their rights, and to respect the rights of others; (ii) encourage community collective action towards shared goals; and (iii) foster non-violent dialogue and conflict resolution. Demand for the training outstripped available funding, and so CEP communities were randomly selected from a larger pool of eligible towns and villages, creating a random treatment and control group for comparison. As part of our evaluation, we surveyed and compared outcomes in the treatment and control communities before and after the program. For results of that evaluation see: Christopher Blattman, Alexandra Hartman, and Robert Blair, "Can we teach peace and conflict resolution? Results from a randomized evaluation of the Community Empowerment Program (CEP) in Liberia " (New Haven, CT: Innovations for Poverty Action, 2011), at http://www.poverty-action.org/project/0139.

[2] In some cases a leader surveyed in Round 1 had been replaced by the time we returned for Round 2. In these cases, we interviewed whoever had assumed the position of the replaced leader.

[3] See Robert Blair, Christopher Blattman, and Alexandra Hartman, "Patterns of Conflict and Cooperation in Liberia (Part 1): Results from a Longitudinal Study," (New Haven, CT: Yale University & Innovations for Poverty Action, 2011). For a nationally representative cross sectional analysis, also see: Patrick Vinck, Phuong Pham, and Tino Kreutzer, "Talking Peace: A Population-Based Survey on Attitudes About Security, Dispute Resolution, and Post-Conflict Reconstruction in Liberia," (Berkeley: UC Berkeley Human Rights Center, 2011).

- Violent strikes or protests
- Violent confrontations between tribes
- Murders
- Rapes
- Beatings or killings of suspected witches
- Trials by ordeal (e.g. "sassywood")

While this measure is obviously imprecise, we believe it accurately reflects the multi-dimensional nature of many conflicts in rural Liberia. Over the past several months we have conducted a series of in-depth qualitative investigating on reports of collective violence gleaned from the survey. These interviews suggest that very few incidents can be neatly classified into the categories above. A murder may incite a protest, which may pit members of different tribes against one another in outbreaks of mob violence. *Sassywood* may be used to adjudicate allegations of witchcraft, which may in turn be related to incidents of sexual assault. A single indicator allows us to capture these multiple dimensions while maintaining our focus on the most destabilizing events.[4]

4. Our methodology

These data provide a unique opportunity to put at least two early warning requirements to the test: whether, from the universe of potential risk factors, it is possible to identify a subset of the most robust correlates of conflict, and whether those correlates can be used to predict conflict more accurately than chance, even in a single time period and a limited number of communities.

The methods we use to answer these questions are complex. Our goal in this report is to highlight a set of easily interpretable, policy-relevant findings, while keeping the discussion as non-technical as possible. We leave the more technical details to a forthcoming academic paper that builds on these results.

Nevertheless, some familiarity with the methods we use will help in interpreting our findings. Throughout this report we rely heavily on three statistical techniques, known as CART, LASSO, and neural networks.[5] These techniques reflect three different approaches to early warning and, more broadly, three different notions of what it means for a community to be "at risk" of conflict.

A. The "CART" method

CART is a diagnostic approach to early warning that treats the risk factors for conflict the same way physicians treat the symptoms of disease. Many symptoms are hierarchical in the sense that doctors use primary symptoms to make quick, preliminary diagnoses, then use secondary and

[4] To attempt to minimize measurement error and reporting bias, we code conflict as having occurred only if two or more of the four leaders surveyed in any given community agree that a given incident did in fact happen.

[5] For readers interested in further detail, a concise review of all three techniques can be found in Trevor Hastie, Robert Tibshirani and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer, 2001.

tertiary symptoms to confirm hunches drawn from primary ones.[6] CART works in precisely this way. It searches first for a primary risk factor that most efficiently distinguishes communities that are most and least to suffer an episode of violence. It then identifies secondary and tertiary risk factors, each of which improves the reliability of the diagnosis.

B. The "LASSO" method

Rather than search for symptoms, the LASSO method attempts to construct a model of conflict in which each risk factor is assumed to have a discrete, additive effect on the probability that conflict will occur. A useful analogy is to the more basic models that climatologists use to predict hurricanes and other natural disasters. The "risk factors" are atmospheric and oceanic conditions—temperature, wind speed and direction, humidity, etc. As any one of these conditions changes, the probability that a hurricane will occur changes with it. The model is adjusted through trial and error, and different risk factors are "weighted" depending on their importance. LASSO works the same way: it assigns each risk factor a weight to indicate how well it predicts conflict. Those that perform well are kept, and those that perform poorly are discarded. Each risk factor is assumed to be independent of all the others, and is either in or out of the model.

C. The "neural networks" method

Unlike CART and LASSO, neural networks treat risk factors as interrelated in complex and unobservable ways. Risk factors are weighted to capture the strength of their *interconnections*, not their independent effects. The name of the technique is an analogy to the structure of the human brain: neurons stimulate responses in the body through the intricate web of signals they send to one another. Operating independently, each neuron plays only a marginal role; it is the accumulation of interactions that matters. Unlike in LASSO, where the effects of each risk factor are linear and additive, here the effects are non-linear and interactive—often massively so.

D. Is any single method better?

Why do we use three methods? Why not just pick the most reliable and stick with it? We believe each of these techniques offers a potentially viable approximation to the way conflict "works" in the real world. Models of this sort are inherently probabilistic, and all methods have their drawbacks. By triangulating across three models, we are better able to identify and manage their relative strengths and weaknesses.

CART yields simple, actionable predictions, but at the cost of over-simplification; neural networks are more nuanced, but often at the expense of clarity and interpretability. LASSO falls somewhere in between. One advantage of CART and LASSO is that each method produces a "short list" of primary risk factors, possibly a simpler tool for policymakers to understand and use. But neural networks may capture important interactions between these variables that the

[6] For example, when emergency responders attempt to identify a heart attack victim, they may ask first about pain in the chest, *then* shortness of breath, *then* pain in other parts of the body. The primary symptom (pain in the chest) is the most efficient one: it allows responders to make as quick and accurate a diagnosis as possible with as little information as possible. Secondary (shortness of breath) and tertiary (pain in other parts of the body) symptoms are used to confirm that diagnosis.

other methods cannot. Rather than rely on one model from the outset, we prefer to test them all and let the data decide.

We begin by assessing the accuracy of each method by looking backwards, using potential risk factors measured in 2008 to predict conflict in 2010. Each model generates a predicted probability that conflict will occur for each of the 247 communities in our sample. We then compare those predicted probabilities to the reported incidence of violence. Our accuracy rates capture the degree of correspondence between what the model predicts will occur and what we know actually happened.

Early warning, however, is not a matter of "retro-casting" into the past, but rather of forecasting into the future. While there is no way to gauge our model's viability for forecasting without collecting new data, we are able to simulate such a test statistically through a procedure known as cross-validation. Cross-validation involves dividing our sample of 242 communities into ten randomly-selected subsets, training the model on some of those subsets, and then testing it on the others.[7] This method has been shown to approximate out-of-sample accuracy rates—the rates we would expect if we assessed the model on new data.

Beyond accuracy, we also attempt to assess the flexibility of each method in balancing "true" and "false" positives and negatives. The costs of conflict are often enormous: lives are lost, property is destroyed, and delicate processes of peace and reconciliation are strained. We believe it is preferable to predict that conflict will occur even when it doesn't than to predict it will *not* occur even when it does. This means that "false negatives" (the model predicts no conflict when conflict actually occurs) are more costly than "false positives" (the model predicts conflict but conflict doesn't occur). There is, however, a trade-off here: to forecast a rare event like conflict, more true positives inevitably means more false positives and lower accuracy overall.[8] We want our model to be flexible enough that it can be adjusted to take this trade-off into account.

5. Results

We began this exercise skeptical that forecasting would yield promising results. Conflict is by its nature idiosyncratic and thus difficult to predict. To our surprise, however, we find that a simple statistical model can retrospectively predict as many as three-quarters of all conflicts measured in the survey while maintaining overall accuracy rates of over 50%. This does not necessarily imply that conflict can be forecasted into the future, but it does get us over the first hurdle for early warning: assessing whether there is any predictable pattern at all.

[7] This exercise required that we reduce the number of communities in our sample from 247 to 242, dropping communities that either were too small (fewer than 20 residents) or had been excluded from the endline CEP analysis. We do not believe these omissions are likely to bias our results in any systematic way.

[8] To see this, it helps to walk through a simple heuristic exercise. Imagine that we have a sample of 100 communities. We know that a riot will occur in one of them, and we decide to try to predict where it will occur. If we want to maximize the overall accuracy of our prediction, our best bet is to predict no conflict at all: our prediction will be, overall, 99% accurate. The problem is that the 1%—the riot—is what we really care about. Because forecasting is by its nature probabilistic, we are unlikely to build a model that predicts all 99 "true negatives" *and* the one "true positive." If we want to predict that 1%, we will have to sacrifice some of our overall accuracy to do it.

A. A simple model can predict conflict with relatively high accuracy rates

Table 1 presents predictive accuracy rates for our CART, LASSO and neural networks (NN) models of conflict. We present our cross-validated results first to approximate the potential for out-of-sample prediction, followed by in-sample results using the full dataset. For each model, we look at overall accuracy rates and rates broken down by type (true positives, true negatives, false positives, and false negatives), as well as the ratio of false positive and negatives to true positives. In our preferred models, false negatives are treated as especially costly such that we err on the side of over-predicting conflict while attempting to maintain overall accuracy rates above 50%.

The LASSO method performs best under cross-validation. Using only a handful of variables measured in 2008, the model predicts almost three-quarters of all outbreaks of violence in 2010 while maintaining an overall accuracy rate of 58%. The ratio of false positives to true positives is 2.9, meaning that for every correct prediction of conflict, the model generates about three "false alarms." This ratio is high, but perhaps not intolerably so given the severe social and economic costs of violence. The ratio of false negatives to true positives is 0.4—impressively low. In theory, the number of false negatives could be reduced even further, but at the cost of overall lower accuracy (and especially more false positives).

Not surprisingly, the performance of all of our models improves dramatically in the full sample. The improvement is starkest for neural networks, which tend to perform much better than simpler alternatives in sample, but much worse out of sample. As in all types of forecasting, the more complex the model, the more it will account for the nuances of a particular set of conflicts, but the less it will generalize to other incidents in other times and places.

Table 2 compares alternative CART, LASSO and neural networks models, varying the extent to which we train them to avoid false negatives. In the first column, false negatives are penalized as especially costly: each model is trained to have lower tolerance for predicting peace where, in reality, conflict occurs. The second column lists accuracy rates under higher tolerance for false negatives. The table displays the absolute number of true and false positive and negatives in each category, as well as accuracy rates and ratios.

For each model in Table 2, as we move from the first to the second column we see that overall accuracy increases while "true positive" accuracy declines. These declines illustrate the tradeoffs implicit in any attempt to forecast a rare event: more true positives means more false positives and lower overall accuracy. Forecasting is a careful balancing act, and the right balance is not easy to find. The point to take away, however, is that even across very different balances, the models perform relatively well.

Table 1: Prediction accuracy for preferred CART, LASSO and NN models

	Cross Validated Predictions			Full Sample Predictions		
	CART	**LASSO**	**NN**	**CART**	**LASSO**	**NN**
Accuracy	53%	58%	52%	78%	64%	67%
True + accuracy	44%	74%	55%	93%	90%	98%
True - accuracy	55%	55%	51%	75%	59%	60%
False + rate	45%	45%	49%	25%	42%	40%
False - rate	56%	26%	45%	7%	10%	2%
Ratio false + to true +	5.0	2.9	4.3	1.2	2.2	2.0
Ratio false - to true +	1.4	0.4	0.9	0.1	0.1	0.0

Table 2: Prediction accuracy for alternative CART, LASSO and NN models

	Cross Validated Predictions					
	CART		**LASSO**		**NN**	
Penalty for false -	**High**	**Low**	**High**	**Low**	**High**	**Low**
True + (of 42)	29	18	31	20	23	11
True - (of 200)	79	111	110	159	102	151
False +	121	89	90	41	98	49
False -	13	24	11	22	19	31
Accuracy	45%	53%	58%	74%	52%	67%
True + accuracy	69%	44%	74%	48%	55%	76%
True - accuracy	40%	55%	55%	80%	51%	25%
False + rate	60%	45%	45%	21%	49%	25%
False - rate	31%	56%	26%	52%	45%	74%
Ratio false + to true +	4.2	5.0	2.9	2.1	4.3	4.7
Ratio false - to true +	0.5	1.4	0.4	1.1	0.9	3.0

B. Our model outperforms intuition-based alternatives, but not by much

To establish a benchmark against which to compare these more sophisticated models, we also attempt to predict conflict using one of several simple decision rules. First, we predict that conflict will recur in 2010 in all the places where it occurred in 2008. In other words, we assume a rate of recurrence of 100%. Second, we predict that conflict will occur only in the most accessible communities—those on or within 15 minutes of a usable road. These models are designed to simulate the sorts of decision-making that we believe may guide early responders in Liberia in the absence of systematic data. Where resources are scarce and data unavailable or unreliable, recurrence and accessibility may be sensible "rules of thumb" for early responders to follow.

Results from these benchmark models are compiled in Table 3 below. Overall they perform surprisingly well—in some cases better (or at least not much worse) than the more sophisticated models described above. Most striking are the predictions based on accessibility (column 3). Nearly two-thirds of all conflicts in 2010 occurred in communities located 15 minutes or less from a usable road. A model based on this variable alone predicts 62% of true positives while maintaining an overall accuracy rate of 54%.

Table 3: Prediction accuracy for benchmark models

	PRIOR CONFLICT	ON ROAD	CLOSE TO ROAD
Accuracy	62%	70%	54%
True + accuracy	48%	29%	62%
True - accuracy	65%	79%	52%
False + rate	35%	21%	48%
False - rate	52%	71%	38%
Ratio false + to true +	3.5	3.4	3.7
Ratio false - to true +	1.1	2.5	0.6

While this "close to road" model yields many false alarms—95 in total—it nevertheless outperforms some specifications of neural networks and LASSO. This is potentially good news for police and the UN, as it suggests that targeting accessible communities may be an efficient way to allocate scarce peacekeeping resources. The other two benchmark models are more accurate overall, but at the expense of high false negative rates—52% and 71% for the "prior conflict" and "on road" models, respectively. Here, again, the trade-off is clear: the "on road" model is the most accurate over all, but also the least likely to predict conflict when it occurs.

C. From 60 potential risk factors we isolate a small number of robust predictors

Which risk factors are the most robust predictors of conflict, and are they consistent across models? In total we coded variables for 60 potential risk factors, all measured at the community level. Rather than select from these 60 on the basis of intuition alone, we allowed our models to identify the most robust correlates through an iterative process of statistical trial and error. The results here are promising as well.

In each model, we find that a relatively small number of variables—often just five to ten—account for the bulk of our predictive power. LASSO and neural networks in particular proved surprisingly effective at isolating a manageable number of correlates from a long list of potential risk factors. CART is by its nature less helpful for weighing the effects of different variables against one another. We are currently exploring an extension of CART's classification algorithm that yields rankings similar to those produced by LASSO and neural networks.

Table 4 lists the 16 most influential risk factors out of the 60 that we tested. A risk factor is included if it appears within the top seven for either LASSO or neural networks. We select variables for inclusion in this list based not on the direction of their effects on the predicted probability of conflict, but rather on their magnitude. In other words, a variable that greatly exacerbates the risk of conflict will rank about the same as one that greatly mitigates that risk, even though their effects work in opposite directions. In cases where the magnitude of the effect was the same for two variables within any given method, we assign those to variables the same rank.

Table 4: Ordinal ranking of risk factors

Variable (measured in 2008)	**Lasso rank**	**NN rank**
Minority tribe has representation in town leadership	1	1
Town population	2	
Proportion of town landless		2
Proportion of town farmers		2
Proportion of town reporting burglary or armed robbery	3	
Proportion of town who contribute to public facilities	4	4
Number of tribes in town (survey-based)	5	
Proportion of town ex-combatants		5
Proportion of town reporting loss of land during war	6	
Degree of victimization by wartime violence in town		6
Proportion of town that believes other tribes are violent		6
Communal wealth index	7	
Proportion of town Muslim	8	
Proportion of town that believes other tribes are "dirty"		8
Proportion of town in largest tribe (leader report)	9	
Proportion of town "strangers"		9
Degree of participation in wartime violence in town		9

While there is some overlap in the risk factors selected by each model, the overlap is incomplete, and in general different methods yield different rankings. Despite these inconsistencies, most of the robust correlates listed above can be sorted into four conceptual categories of risk: population size, ethnic and religious diversity, presence of ex-combatants and exposure to wartime violence, and communal wealth.

1. Population size

Of the 60 variables we tested, town or village population is among the best predictors of conflict. In a sense this result is merely mechanical: more people means more potential disputants, more potential victims and more potential perpetrators of crime. Nevertheless, the relationship is striking. Figure 1 estimates the likelihood of conflict in 2010 as a function of town population in 2008.[9] The solid line is the predicted probability that conflict will occur, and the gap between the two dashed lines capture the degree of uncertainty in that prediction (the wider the gap, the greater the uncertainty). The relationship is almost linear: for every additional 1000 residents in a given town or village, the predicted probability of conflict increases by approximately 8 percentage points on average.

This finding is consistent across two of our three models. Only one other variable has such a strong and positive effect on the predicted probability of conflict in LASSO. Using CART we see that of the 200 communities in our sample with populations under ~4000, the probability of conflict is just 14%; among the 42 communities with populations over ~4000, the probability is almost 50%. Looking at populations that fall above and below this threshold is the single most efficient way to distinguish high risk from low risk communities.

2. Ethnic and religious heterogeneity

Our measures of ethnic and religious heterogeneity are robust predictors of conflict as well. We explored this relationship in a previous policy report,[10] and our more rigorous statistical analyses confirm these initial findings. The relationship is in some ways complex, but in general the risk of conflict is lower in communities populated predominately or exclusively by a single tribe.

We constructed several indicators of ethnic and religious heterogeneity, but focus here on two of those measures. First, we counted the number of different tribes represented among the 20 randomly-selected respondents in each community (ranging from one to seven). Second, we calculated the proportion of those 20 respondents belonging to the largest tribe in town (ranging from under 40% in the most heterogeneous communities to 100% in the most homogeneous ones). Third, we asked local leaders to estimate the population of their communities, then asked them to estimate the size of various sub-populations, including ethnic and religious minorities and "strangers" or new migrants.

Figures 2a and 2b below capture changes in the predicted probability of conflict as each of these first two measures of heterogeneity increases. The relationship is linear and positive: the more heterogeneous the community in 2008, the greater the likelihood that conflict will occur in 2010. This finding is consistent across our other measures of heterogeneity as well.

Note that these figures do not tell us which tribes are most likely to perpetrate, or be victims of, violence. Nor, in fact, do they tell us whether conflict occurs within or between tribes in these

[9] Again keeping the discussion as non-technical as possible, the figure is derived from a "smoothing" function in which separate bivariate regressions are run across many small segments (or "bandwidths") of the data.

[10] See Blair, Blattman, and Hartman, "Patterns of Conflict and Cooperation in Liberia (Part 1): Results from a Longitudinal Study."

more heterogeneous communities. It may be that most violence occurs among members of the same tribe even in the most diverse towns and villages. We cannot disentangle these possibilities using our data alone. These are important caveats, and suggest fruitful avenues for future data collection and analysis.

What explains this relationship between diversity and conflict? Several interpretations strike us as plausible. One possibility is that inter-tribal tensions continue to simmer where multiple tribes cohabitate, occasionally erupting into violence. This is consistent with the commonly held belief that inter-tribal rivalries remain an important and recurring cause of conflict in Liberia.

Another possibility is that institutions tend to be more exclusionary in more diverse communities, and that exclusion foments conflict between dominant and marginalized tribes. Our results, however, seem to contradict this hypothesis. We find that the risk of conflict is *lower*, not higher, where the minority tribe is excluded from positions of local leadership—a relationship that holds across both more and less diverse communities. Where ethnic minorities are represented in local government, the predicted probability of conflict is 33%; where they are not, the probability drops to just 13%. This result is puzzling, but may suggest that while power sharing and political mobilization helps minorities articulate their demands to and within government, mobilization also helps them organize for violence when those demands are not met. This is mere speculation, however, and other interpretations are possible.

Yet another possibility is that our results are spurious. We find, for instance, that the risk of violence is highest in the largest communities. If the largest communities also tend to be the most diverse, then what appears to be a correlation between diversity and conflict may in fact be a misleading artifact of the relationship between conflict and population size. This interpretation strikes us as unlikely, however, as the correlation between heterogeneity and violence holds even after we control for population. Finally, it is possible that there is some "omitted variable" that causes both conflict and diversity, and that we have failed to include in our model. We cannot rule out this possibility. Nevertheless, while heterogeneity may not *cause* conflict, it certainly helps us predict where conflict is most likely to occur.

3. Presence of ex-combatants and exposure to wartime violence

Ex-combatants and legacies of wartime violence continue to shape conflict dynamics in rural Liberia. Two of our three models identify the presence of ex-combatants as a robust predictor of violence. The important distinction here is not between communities with many ex-combatants and those with few, but rather between those with any ex-combatants and those with none. CART sets its diagnostic threshold at close to zero: where the per capita population of ex-combatants is lower than 0.3%, the predicted probability of conflict is 8%; where the population is higher than 0.3%, the predicted probability is 20%.

This does not imply that ex-combatants are necessarily the perpetrators of violence; indeed, as we noted in our previous report, ex-combatants may sometimes act as de facto law enforcers where the police are absent or ill-equipped. We interpret these findings more broadly to suggest that experiences of wartime violence may continue to foment tensions even in peacetime.

Figure 1: Relationship between population and conflict risk

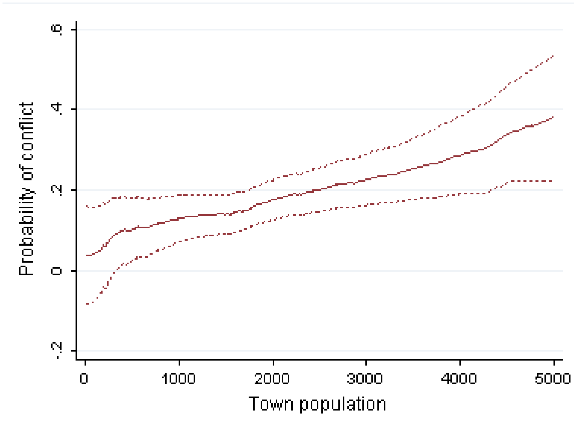


Figure 2a: Relationship between heterogeneity and conflict risk

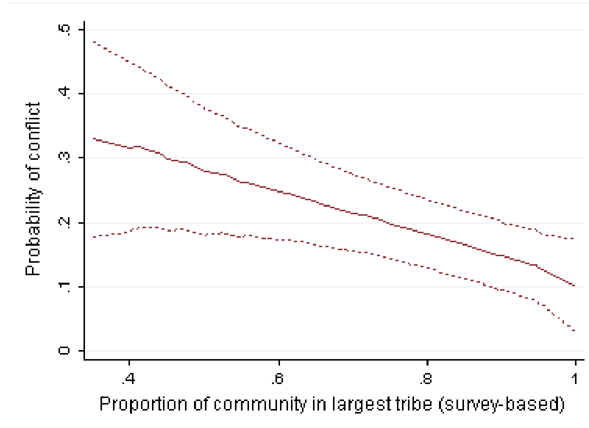


Figure 2b: Relationship between heterogeneity and conflict risk

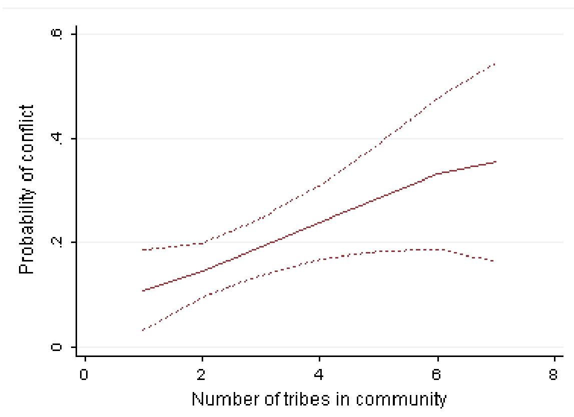


Figure 3a: Relationship between wartime violence and conflict risk

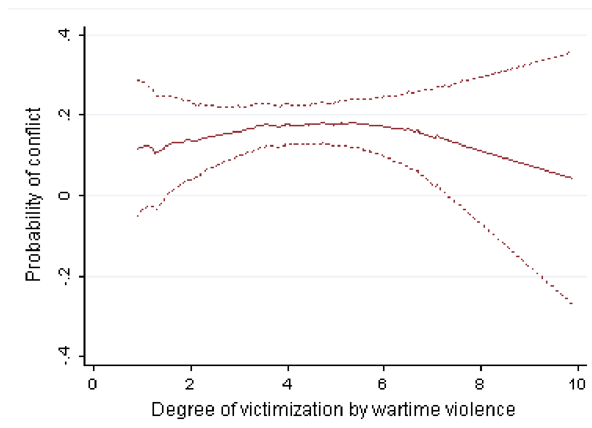


Figure 3b: Relationship between ex-combatants and conflict risk

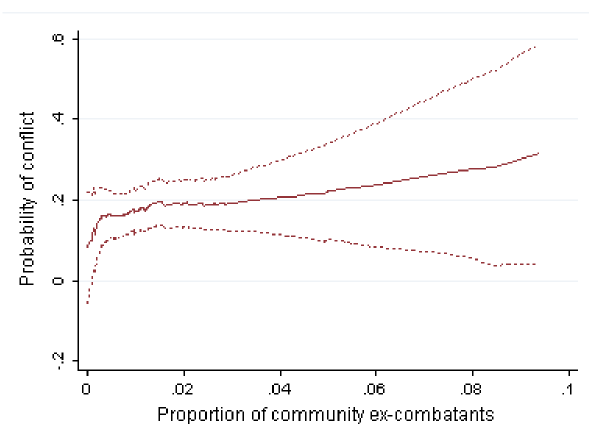
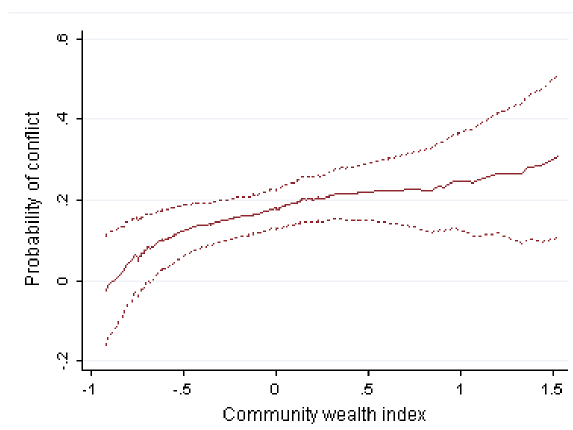


Figure 4: Relationship between wealth and conflict risk



To measure these experiences more generally, we asked community members about their exposure to 14 different types of wartime violence. We then construct two community-level indices to aggregate those experiences across the 20 randomly-selected community members in each town or village: one for victimization and another for participation in violence. Our neural networks model selects both of these indices as robust correlates of conflict, though the relationship between victimization and violence appears to be non-linear and more uncertain at higher levels of exposure (Figure 3a). Loss of land during the war is predictive as well: in communities where more than 8% of residents lost their land, the risk of conflict is 25%; in communities where fewer than 8% lost their land, the risk is only 11%.

4. Wealth and assets

The threat of conflict is higher in relatively rich communities compared to relatively poor ones. To measure wealth, we asked respondents a variety of questions about individual and household assets; we then construct an index of intra-communal wealth by averaging across those individual responses. The effect of wealth on the probability of conflict is positive and more or less linear, though it is most pronounced on the left (poorer) side of the distribution (Figure 4). The poorest communities are the least conflicted, but a slight increase in wealth predicts a dramatic increase in the risk of conflict at this relatively poor end of the spectrum.

We also find that the predicted probability of violence is lower where majorities of residents work in farming rather than other occupations (e.g., small business and trade). This likely reflects the disparities in wealth between predominately agricultural and predominately commercial communities. Consistent with the correlation between wealth and conflict, the likelihood of violence is lower in these poorer, more agricultural communities.

These findings seems to confirm the intuition that conflict is more likely where there are more resources to dispute, but seems counter to the conventional wisdom that poor communities are less secure and more vulnerable to conflict than wealthy ones. Also, while poverty in most assets seems to predict a decline in the risk of conflict, poverty in *land* does not. Higher rates of landlessness predict higher rates of violence, and neural networks select landlessness as the second most influential risk factor out of 60 candidates.

Of course, not all of the variables listed in Table 4 fit tidily into one of these four categories. We find, for example, that the higher the proportion of respondents that report a burglary or armed robbery in 2008, the higher the likelihood that their community will suffer a larger-scale incident of violence in 2010. This is intuitive, implying that communities prone to certain types of crime and violence tend to be afflicted by other, potentially more serious conflicts as well.

Not all our results are so intuitive, however. We find, for example, that the larger the proportion of respondents who report contributing labor or cash to public facilities in a given community, the greater the predicted probability of conflict. This runs to counter to our expectation, and to the conventional wisdom that more socially cohesive communities—those more capable of intra-communal collective action—should be less prone to violence. There are many potential explanations for this finding. For now, we leave it as a puzzle pending further analysis.

These anomalies notwithstanding, most of our robust risk factors seem to fit within this simple conceptual scheme, and together yield predictions that are far more accurate than we expected. These robust risk factors also tend to be inter-correlated, constituting, perhaps, a sort of "syndrome" of conflict risk. In Liberia, like in most countries, big urban communities tend to be more accessible and more ethnically and religiously diverse, and usually attract larger concentrations of wealth and business than their small, rural counterparts. These communities may also have been targeted by armed groups during the war as especially desirable loci of territorial control. While we are hesitant to make specific predictions about where conflict is most likely to occur in the future, we believe this "syndrome" perspective may offer general but useful guidance to inform funding and programming decisions within the EWER community in Liberia.

6. Conclusions and implications

A. Forecasting looks more promising than expected

Our results suggest that the prospects for conflict early warning and forecasting may be more promising that we first imagined. That we are able to predict conflict with relatively high accuracy rates even in our out-of-sample simulations suggests that true forecasting, rather than mere "retro-casting," may be a viable possibility. That so many of our robust correlates are easy to measure suggests that the costs of testing our model against new data may not be prohibitively high.

These results come with many caveats. First, our model is built on just two rounds of data covering three counties over two years—hardly enough to judge its capabilities for real-world EWER. Second, our forecasts are retrospective rather than prospective, and there is no way to know how the model will perform if projected into the future.

Third, all of our potential risk factors are measured at a single moment in time. This allows us to assess the effects of their *levels* (e.g., population of ex-combatants in a given community) but not *trends* (e.g., an influx of ex-combatants into a given community) on the probability that conflict will occur. Trends, however, may be more important than levels for purposes of forecasting. For example, a community that is and has always been poor may be less likely to experience violence than one that suddenly becomes poor. Assessing this latter possibility would require further data collection to identify trends over multiple periods of time.

All of these limitations should be kept in mind in interpreting our results. Caveats notwithstanding, however, we believe our results constitute a promising first step towards data-driven EWER in Liberia.

B. Early responders may be distributing resources more or less efficiently already

What sorts of communities do police, peacekeepers and NGOs usually target? Where do most peacebuilding interventions occur? Much programming tends to cluster in larger, more accessible communities with sufficient infrastructure to support field offices and staff. (Indeed, a common critique of peacebuilding programs in Liberia is that they focus too much on urban centers and not enough on rural towns and villages.) Outbreaks of violence also tend to attract persistent attention from government and NGOs, and "conflict-proneness" in the future is often equated with histories of conflict in the recent past.

Is this approach misguided? Or have policymakers succeeded in channeling scarce resources to the areas that need them most? In previous reports we have criticized the tendency to over-program a relatively small subset of counties and communities, especially those that garner disproportionate media attention (Voinjama is an obvious case in point here). While we believe that critique remains valid, our results here suggest that policymakers may in many cases be distributing resources efficiently already, without the benefit of whatever insights a more data-driven risk assessment might provide.

We find that a large proportion of conflicts tend to occur in diverse, heavily populated, reasonably accessible communities that suffered bouts of violence either during the war or in the more recent past. If police and peacekeepers tend to intervene most often in these places as well, then our results may bode well for existing approaches to early response. We also suspect, however, that there are many towns and villages in rural Liberia that fit this profile for risk but that nevertheless get ignored. Existing modes of conflict prevention and mitigation may be well-targeted, but there is almost certainly room for improvement.

C. Next steps for a data-driven early warning system

Our analysis raises a number of questions for data-driven EWER that we believe merit further exploration. How well would our models perform if we tested their predictions against future events that are yet to occur, rather than past events we have already observed and measured? How well would they perform if we attempted to extrapolate beyond our sample to other counties and communities? Will the same risk factors that help us "retrocast" conflict in 2010 remain robust predictors in 2012 and beyond? And if so, can we measure those risk factors at higher frequency to generate shorter-term (or even real-time) predictions?

To answer these questions, we envision three potential next steps:

1. *Test the generalizability of the model to 2011 and 2012.*

One option is to continue to validate our model with another out-of-sample test using new, quickly-gathered data. For example, we can replicate the forecasting exercise described above using our existing data on risk factors in 2008 and new data on conflicts in 2011. We estimate this to be the fastest and least expensive means of testing the generalizability of our models, requiring only a re-survey of the same local leaders in 2012 to ask about incidents that occurred over the past year.

2. *Test the generalizability of the model to a nationally representative sample.*

To increase our confidence in the model's out-of-sample stability, we can test its predictions on a new, perhaps nationally-representative sample of communities. One possibility is to conduct additional data collection in a subset of communities surveyed by the Human Rights Center at Berkeley in 2010. This would allow us to update our risk factors, incorporate trends as well as levels, and assess the model's generalizability beyond Lofa, Nimba and Grand Gedeh counties.

3. *Pilot a simple high-frequency system of conflict forecasting.*

Finally, we can design a system that tracks trends in the most robust correlates of conflict at higher frequency using, for example, Ushahidi's SMS crowd-sourcing platform. In the short term, high-frequency data would allow us to train and test our model against many successive rounds of data collection; in the longer term, it might yield actionable predictions for real-time EWER. Whether such a system is even logistically and technologically feasible remains an open question, but our results suggest it may be a question worth answering.

D. General implications for the early warning community

The above is only a blueprint research agenda, one that focuses rather narrowly on the data-driven approach to EWER. This approach is by no means the only one. We believe there are several lessons to be drawn from our results that may be applicable beyond the world of statistical analysis.

First, we urge the EWER community in Liberia to focus less exclusively on analysis of past conflicts and more on anticipation of future ones. This lesson is especially relevant to the members of the Early Warning Working Group (EWWG). The EWWG has made important contributions to the analysis of conflict in Liberia by insisting on data quality and the verification of incident reports. Much like our own statistical models, however, the Working Group's analyses thus far have been mostly backward-looking. Analyzing past incidents is essential for predicting future ones, especially where violence is cyclical or where underlying tensions and grievances languish under temporary fixes. Yet rarely do organizations in Liberia project beyond these retrospective analyses. With the possible exceptions of UNMIL and the International Crisis Group, very few attempt a more forward-looking approach to early warning.

The EWWG has a wealth of knowledge and ear-to-the-ground experience on conflict dynamics, and is well positioned to fill this gap. Forecasting need not involve specific predictions about where and when conflict will occur, but it does require looking ahead and developing expectations about the future rather than simply looking behind into the past.

This raises a second point. Statistical EWER sometimes perpetuates the misconception that forecasting demands data, preferably the quantitative kind. This is wrong. All approaches to early warning hinge on the assumption that conflict is not merely idiosyncratic and can therefore be predicted (in theory at least) with some probability greater than chance. If we believe conflict is random, then the notion of "early warning" is vacuous. How can we hope to "warn" of events that might occur anywhere, anytime, for any reason?

If we believe conflict is *not* random, then we must also believe in probabilistic forecasting. In other words, we must believe in the possibility of analyzing past conflicts, identifying salient risk factors, and using those risk factors to anticipate where conflict is more or less likely to occur in the future.

At no point does this process need to involve statistics. Data-driven risk assessments such as ours stand to benefit greatly from first-hand, qualitative knowledge and intuitions about trends on the ground. These two types of analyses should be thought of as complements rather than substitutes for one another. Forward- and backward-looking analyses are complementary as well. Attempting prediction forces us to sharpen our understanding of past conflicts and more explicitly articulate our intuitions about the drivers of violence. This is a useful exercise for all approaches to early warning, data-driven or otherwise.

To think about conflict in this way may strike some readers as counterintuitive. After all, engaging in violence is a choice, and that choice is sometimes made at surprising times and for surprising reasons. Conflict, in this sense, seems neither predictable nor random. If this view is right, then there is little hope for EWER.

We believe this understanding of violence is accurate, but only partly so. Our analysis says nothing about the motivations for violence, and focuses instead on the structural or "ecological" conditions that seem to make violence more likely. These conditions do not cause violence, but they do seem to heighten the risk that people will make the choice to engage in violence. Motivations that cannot be predicted or understood often produce patterns of conflict that can.

Third, we encourage the EWER community to consider standardizing at least some of its indicators for conflict and risk. Members of the EWWG have expressed interest in standardization in the past; we believe this remains an important and unfulfilled goal for early warning efforts in Liberia.

Again, statistics are not necessary—if anything, this is more an analytical exercise than a statistical one. To give but one example, the slow pace of "reconciliation" is often cited as a catalyst for conflict. In her Nobel Peace Prize acceptance speech, President Ellen Johnson Sirleaf concluded from the violence preceding her reelection that "we have a whole group of people from 1989 till now who haven't been reconciled," and that "the need for reconciliation is great." But what does this mean? How can we tell if someone is or is not "reconciled?" Even if reconciliation is not measurable, is there at least some way we can know it when we see it?

The same questions could be asked of countless other potential risk factors for conflict: poverty, inequality, exclusion, infrastructure, institutions, etc. Too often risk factors such as these get recited as though their meanings were self-evident. The EWER community can help correct this problem by coming to some consensus on what these risk factors are, what they mean, and how they can be monitored or even measured.

The EWER community should also beware of undiscriminating "laundry lists" of risk factors. Recitations of all the potential variables that may exacerbate the risk of conflict will be unhelpful without some way to distinguish strong predictors from weak ones. The method we use relies on statistical algorithms, but other options exist and may in many cases be preferable to our mechanical approach. Whatever method is chosen, some effort must be made to winnow down these laundry lists if they are to yield usable insights.

Finally, combining these first three points, we urge the EWER community to direct attention to places where the risk factors for conflict have materialized but conflict itself has not. If, for example, we believe our finding that ethnically and religiously diverse communities tend to be at greater risk of violence, then we should devote at least some energy and resources to heterogeneous communities where violence is rare.

The reasoning behind this recommendation is two-fold. First, by targeting high-risk but low-conflict communities, it may be possible to prevent violence before it occurs. Second, by learning about the dynamics of conflict and risk in these places, we may better understand how some communities successfully manage the threat of escalation. This is an aspect of early warning that is often overlooked. Understanding why conflict does *not* occur, even when we think it will, is equally if not more important than understanding why it does.

E. Implications for targeting of security and peace-building services

Forecasting conflict is only one element in the calculation that policymakers must make about how or whether to respond. Allocations of scarce resources should depend on evidence that those resources will be effective at mitigating or preventing conflict. Forecasting is only useful if we also have a sense of what forms of interventions are effective, and why.

Our dataset was constructed around the evaluation of one of Liberia's flagship peacebuilding projects: the Community Empowerment Program (CEP). We might ask, what if the CEP had been targeted to the most conflict-prone communities as identified by our model? Would we have expected to see disproportionately larger impacts, and does this imply that our forecasts can be used to target assistance?

The short answer is "maybe." The trouble with answering this question in this context is that the CEP had very little impact on large-scale conflict. Indeed, as the evaluation report highlights, overall levels of conflict increased rather than decreased in communities that (randomly) received the program, in large part, we think, because the training stimulated discussion and possibly opened old wounds. A lower *proportion* of these conflicts were violent than in communities that received the CEP, but since the absolute number of conflicts was higher, the two effects (higher conflict, lower violent conflict) roughly cancel one another out, so that levels of violent conflict (as defined in this report, for predictive purposes) do not change appreciably.

As a consequence, it is no surprise that the CEP was not systematically more or less effective in high-risk communities. This "non-finding" has an important lesson: future peacebuilding and security interventions ought to be evaluated for their effectiveness, and the value of a conflict forecasting system ought to be predicated on its ability to help target scarce funds and attention more effectively.