TC ABSTRACT

I. Basic Project Data

▪ Country/Region:	REGIONAL/Regional
▪ TC Name:	Development of Predictive Model to build pilot risk tool to identify victims of domestic violence
▪ TC Number:	RG-T3272
▪ Team Leader/Members:	Pantzer, Robert (IFD/ICS) Líder del Equipo; Serrano Berthet, Rodrigo (IFD/ICS) Jefe Alterno del Equipo de Proyecto; Hidalgo, Nidia (SCL/GDI); Pombo Rivera, Cristina (SCL/SCL); Acuna Castillo, Nathyeli Yethzi; Catano Guzman, Mariana (IFD/ICS); Verissimo Da Silva, Carolina; Azevedo, Andrea Pereira (LEG/SGO).
▪ Taxonomy:	Research and Dissemination
▪ Number and name of operation supported by the TC:	N/A
▪ Date of TC Abstract:	22 May 2018
▪ Beneficiary:	Honduras, Uruguay y un estado de Brasil (por definir)
▪ Executing Agency:	Inter-American Development Bank
▪ IDB funding requested:	$ 200,000.00
▪ Local counterpart funding:	$ 0.00
▪ Disbursement period:	18 months
▪ Types of consultants:	Individuals; Firms
▪ Prepared by Unit:	Innovation in Citizen Services Division
▪ Unit of Disbursement Responsibility:	Institutions for Development
▪ TC included in Country Strategy (y/n): ▪ TC included in CPD (y/n):	No No
▪ Alignment to the Update to the Institutional Strategy 2010-2020:	Institutional capacity and rule of law; Gender equality; Diversity

II. Objective and Justification

2.1 Domestic Violence (DV) cannot be deterred by routine patrol in crime “hot spots.” The most serious forms of DV – those that involve serious injury or even death – occur at a tragically high rate in an absolute sense, but account for a small share of the literally hundreds of thousands of DV events that occur in a given jurisdiction each year. Identifying the victims who are truly at greatest risk from among many victims who appear, a priori, to be at risk, is like finding a needle in a haystack for the law enforcement.

2.2 The objective consists in building a world-class prediction model that can assist stakeholders like law enforcement agencies or victim service providers in doing exactly this. Building upon the experience of the Chicago Crime Lab (CCL), the research team would engage with 2-3 jurisdictions in Latin America (cities or central governments) to build a state-of-the-art risk tool to identify victims of domestic violence who are at greatest risk of experiencing repeat victimization. This exercise will produce two outputs: 1) a probability of repeat victimization for each victim known to law enforcement based on that victim’s prior contact with law enforcement and other public agencies and 2) a list of predictors that are most highly correlated with repeat victimization. By efficiently extracting the true “signals” from the large amount of available data, this tool will be helpful to law enforcement in targeting their most promising victim- or offender-based intervention strategies.

2.3 The statistics on DV are grim: According to the World Health Organization (WHO), nearly 1 in 3 women in Latin America will experience violence at the hands of an intimate partner during her lifetime. Domestic violence survivors often aren't the only victims – approximately half of all DV occurs in households in which children under the age of 12 are present. DV imposes a terrible burden, not only on the victim herself, but also on her children. Sadly, this burden compounds itself generation after generation, becoming an engine for the inter-generational transmission of violence.

2.4 While homicide rates in the industrialized world have declined considerably since their peak thirty years ago, DV has remained high. In some cities, the percentage of homicides that are DV-related has more than tripled, with more than half of female homicide victims being killed by an intimate partner. Seeking a proactive approach to reducing violence, police departments and social service agencies have developed a host of interventions that engage with victims to provide critical services and ensure the victim's continued safety. In a world of scarce resources, a key question is: How can the police target scarce resources most effectively?

2.5 The Challenge: While police departments around the world have been enormously successful in using data to reduce street crimes, it has been difficult for this approach to be successful in reducing DV. This for two reasons. First, most of the domestic violence occurs behind closed doors. As a result, domestic violence cannot be deterred by routine patrol in crime "hot spots". Second, the most serious forms of DV – those that involve serious injury or even death – occur at a tragically high rate in an absolute sense, but account for a small share of the literally hundreds of thousands of DV events that occur in a given jurisdiction each year. Identifying the victims who are truly at greatest risk from among many victims who appear, a priori, to be at risk, is like finding a needle in a haystack for the law enforcement.

III. Description of Activities and Outputs

3.1 Is it possible to build a credible data-driven approach to concentrate resources on the most at-risk victims thereby greatly increasing the social good per available dollar of public funding? Using state-of-the-art techniques from data science and Machine Learning (ML), the idea is to build a world-class prediction model that can assist stakeholders like law enforcement agencies or victim service providers in doing exactly this.

3.2 **Component 1. Development of Predictive Model:** Building upon the experience of the Chicago Crime Lab (CCL), the research team would engage with one or two jurisdictions in Latin America (cities or central governments) to build a state-of-the-art risk tool to identify victims of domestic violence who are at greatest risk of experiencing repeat victimization. This exercise will produce two outputs: (i) a probability of repeat victimization for each victim known to law enforcement based on that victim's prior contact with law enforcement and other public agencies; and (ii) a list of predictors that are most highly correlated with repeat victimization. By efficiently extracting the true "signals" from the large amount of available data, this tool will be helpful to law enforcement in targeting their most promising victim- or offender-based intervention strategies. Preliminary results from CCLs work in other jurisdictions indicate that, in predicting an individual's risk of felony re-victimization, the model outperforms human decision-making by at least 30%. CCLs preliminary model likewise appears to be considerably more effective than the field's numerous current risk tools – and, as layers in additional data are added and the model is improved, the effectiveness of this model will continue to increase.

3.3 **Component 2. Testing Intervention in Field:** A challenge that emerges when applying machine learning tools to policy problems, arises when the data available to train and evaluate a new prediction model relative to current human decisions are themselves the product of existing human decisions. In the context of domestic violence, it may that the

status quo regime already identifies high risk victims and applies and intervention to them. A naïve retrospective comparison between machine and human selections could indicate that the algorithm selected individuals are re-victimized at a higher rate than those selected by business-as-usual decision making. But if the intervention is efficacious, then it could be the case that without the intervention, human-selected victims would have been re-victimized at a higher rate. This phenomenon is called "polluted outcomes" and one of the most rigorous procedures for overcoming it is via a randomized control study in the field. The prospective RCT solves this problem since average outcomes of two equivalent groups are compared, where treatment assignment is made either by the status quo targeting system or by the new statistical decision tool. The RCT also addresses the issue that humans may select on the basis of both risk as well as responsivity to treatment, while the algorithm selects only on the basis of risk. The RCT does this by estimating whether machine learning, on net, enhances the treatment effect of the intervention.

IV. Budget

Indicative Budget

Activity/Component	IDB/Fund Funding	Counterpart Funding	Total Funding
Component 1. Development of Predictive Model	$ 100,000.00	$ 0.00	$ 100,000.00
Component 2. Testing Intervention in Field	$ 100,000.00	$ 0.00	$ 100,000.00
Total	$ 200,000.00	$ 0.00	$ 200,000.00

V. Executing Agency and Execution Structure

5.1 The operation will be Bank executed and under the technical supervision of CCL. CCL team includes: Aaron Chalfin, Assistant Professor of Criminology at the University of Pennsylvania and is an affiliated researcher at the University of Chicago Crime Lab. Prior to joining the faculty at Penn, he spent two years as Research Director at the University of Chicago Crime Lab's NYC office. His current research portfolio contains a mix of evaluation research and prediction projects that use machine learning to guide the efficient allocation of scarce criminal justice resources. His past research has considered the effect of police manpower on crime, the relationship between crime and unauthorized immigration, and both the cost and deterrent effect of capital punishment. Aaron holds a Ph.D. from the Goldman School of Public Policy at UC Berkeley.

5.2 Working with Aaron will be Zubin Jelveh, Research Director for Data Science at the University of Chicago Crime Lab's NYC office. Zubin's research focus is computational social science. Specifically, he builds and studies predictive models of human behavior in the domains of criminal justice and academic publishing using techniques from machine learning and natural language processing. Additionally, he studies the – sometimes unintended – consequences of these predictive algorithms. He has also developed record linkage algorithms that are tailored for the unique features of criminal justice data. Zubin holds a Ph.D. in computer science from New York University's Tandon School of Engineering.

VI. Project Risks and Issues

6.1 As with any policy or program that has implications for real people's lives, there are ethical considerations with this proposed Technical Cooperation that must be carefully considered and managed. For example, there might be adverse effects with this (like any) intervention, if for example an intimate partner became angered by police showing

up. We recommend the establishment of an advisory board designed to deliberate on and address potential implementation issues

VII. Environmental and Social Classification

7.1 The ESG classification for this operation is "undefined".