Appendix A

Mexican Labor Law, Description of Data, and the Calculator

The appendix is divided into four sections. The first describes the Mexican labor law and court procedures, the second provides some additional details on the administrative data, the third describes our surveys and response rates, and the fourth describes the calculator we use to make predicted outcomes for the cases in which we intervene.

9.1 Mexican Labor Law

Mexican labor Law: On paper, Mexican labor law is protective of workers. For any employer-initiated separation that is not a ''justified dismissal,'' the law requires severance payment of a minimum of 90 days' wages and benefits. The law provides very few legal bases for a justified dismissal. For example, workers are owed severance if they are fired for low productivity or because of poor market conditions. As an element of their lawsuit, workers may ask the court to order the firm to rehire them. If they win their case and the firm refuses to rehire them, then they are entitled to an additional severance payment equal to 20 days' wages for each year of tenure at the firm.[31]

The Process: In a firing lawsuit, an initial claim is filed by the worker or her lawyer, and an initial hearing date is set, generally two to three months after the date of filing. The defendant(s) must be notified of the filing and hearing date by a formal court summons that must be delivered in person by an employee of the court. The notification process typically takes 6 months.[32] Once notification of all defendants has taken place, a ''conciliation, demands, and answers'' hearing takes place. The principal demand in most firing lawsuits is either the base severance pay of 90 days or reinstatement of employment.

If a settlement cannot be reached then the defendant must answer the lawsuit. Additional hearings are scheduled for presentation and viewing of evidence, after which the written record is closed. All proceedings are conducted by an administrative assistant to the judge. Hearings are transcribed into the case file. The file is passed on to the judge, who writes the final

[31] Certain categories of workers (those defined as ''at-will'') are also entitled to to 20 days' wages per year of tenure. An at-will worker is one who is employed in a position of confidence, for example, a driver or a personal security guard. The law recognizes that the employer may need to dismiss the worker if that confidence is broken.

[32] In practice, notification involves substantial corruption, as the lawsuit cannot proceed without it. Aldeco Leo, Kaplan and Sadka (2014) shows that when notifiers' work load is assigned randomly and control of case files is taken away from notifiers, rates of successful notification more than double.

decision. Enforcement of judgments involving payments to the plaintiff is often challenging. A large proportion of firms do not pay the judgment voluntarily, and a seizure of assets must officially be performed by officers of the court. This is followed by adjudication of liquid assets or sale of non-liquid assets to pay the worker the awarded amount. Given that an interval of six-months between hearings is typical, the fact that there are more than four hearings held in an average case, and the frequent postponement of scheduled hearings due to lack of notification, the average lawsuit continuing to a judge's decision takes over three years. This in spite of the law stipulating that lawsuits should have a maximum duration of 100 days.

Lawyers: Once a case is filed, lawyers control the lawsuits almost completely. The presence of the parties themselves at the hearings is not compulsory, unless they are to be deposed as part of the evidentiary hearings. By law, workers who are not able to hire a private lawyer must be provided with free public legal assistance from labor lawyers in the public prosecutor's office. Public lawyers are paid a flat wage by the court and may not charge clients anything further. A public lawyer handles as many as 400 cases concurrently, while administrative data suggest that a normal load for a private lawyer is at most 50 cases. So while public lawyers are generally well qualified, they have an incentive to finish cases quickly in order to reduce their workload. The general perception is that public lawyers do not use creative or aggressive litigation strategies that may have a positive payoff but imply longer cases or more work.

Private lawyers must be licensed, but obtaining the license is fairly easy and otherwise lawyers are unregulated. Survey data indicate that 82 percent of workers are suing for the first time. Workers have little access to information about where to find a good private lawyer, and often opt for the first one that they run into, with little notion of that lawyer's reputation or previous record. The surveys show that 38 percent of plaintiffs using a private lawyer say that they found the lawyer either just outside the court or in one of the court corridors. Plentiful anecdotal evidence suggests that these "informal lawyers" are low-quality and may not serve their clients' best interests.

Private plaintiff's lawyers typically charge an initial fee of about MXN$2000 pesos (USD 100) to file the lawsuit and a contingency fee of about 30 percent of any amount collected by the plaintiff. In spite of the contingency fee, their incentives are not perfectly aligned to those of their clients. First, while plaintiffs are party to a single case, the lawyers manage a portfolio of many firing lawsuits with widely differing characteristics, against many different firms. With diversified risk, they may be more willing to take risks on any given case. Second, filing a low-quality suit is cheap and easy and the lawyers may profit from collecting the filing fee even with no expectation of recovering anything on behalf of the worker. The plaintiff must ratify settlements of the lawsuit in person, as well as acceptance or rejection of an offer of reinstatement, should the plaintiff's side receive one. Lawyers typically do not bring the plaintiffs to hearings nor do they provide them much detail on the developments in the lawsuit. As will be shown below, the physical presence of the worker at the hearing in which we intervene in the field experiment is crucial for the effectiveness of our intervention.

9.2 Administrative Data

The main source of administrative data are the initial case filings for the historical and experimental cases. For the experimental cases, we also collected data on the phase 1 and 2 case outcomes in December 2018 and in January and July 2020. The historical data are used to develop a predictive model of case outcomes. We then use this model and the data from each experimental case filing to predict outcomes for each case, as described in subsection 9.4 below. Because we intervene in many cases at the time of their first hearing, we use only data from the initial case filing. The case files are available in hard copy, and we digitize the variables shown on Table A.1.

Table A.1: Casefile Variable

VARIABLE	DESCRIPTION
Tenure	Employee's tenure with the employer.
Weekly hours	Number of hours that the plaintiff worked on a weekly basis.
Reinstatement*	The plaintiff claims reinstatement.
Severance pay*	The plaintiff claims constitutional indemnity (three months of integrated salary) that the law dictates for unjustified dismissal.
Lost wages*	The plaintiff claims lost wages/back pay.
Vacation pay*	The plaintiff claims accrued vacation days not taken.
Overtime*	The plaintiff claims overtime pay.
Twenty days compensation*	The plaintiff claims the payment of compensation (20 days per year worked) that the law dictates for unjustified dismissal for a worker who has the right to be reinstated but the employer refuses to reinstate, or for an at-will employee who cannot ask for reinstatement.
Insurance-IMSSIN-FOSAR*	The plaintiff claims the payment of employer contributions that were not made to these institutions, or retroactive registration in the institutions (SAR: retirement savings, IMSS: social security, INFONAVIT: worker's housing fund).
Co-defendant*	At least one of the codefendants is the IMSS or INFONAVIT or SAR.
Total asked	The total quantifiable peso amount of the worker's claim.
Minimum legal entitlement	The quantifiable peso amount of the sum of severance pay, vacation and end of year bonuses of the last year of tenure at the firm. It is a conservative estimate of the minimum amount of money the worker is entitled to if she wins the lawsuit.

Notes: Detailed description of variables used throughout the paper. Dummy variables are marked with *.

9.3 Survey Data

We conducted surveys with participants at the hearings in Phases 1 and 2, and with dismissed workers approaching the court in Phase 3. The surveys in the first two phases had to be very short so that we did not delay the start of hearings. We use one survey instrument for plaintiffs, and another for the lawyers representing both plaintiffs and employers. Table A.2 shows summary statistics for workers and lawyer from phase 1. Table A.3 shows compliance rates for both the surveys and treatments for the first two phases. We have no measure of noncompliance in phase 3. Every worker who completed the survey and was assigned to treatment received the calculator treatment. However, there were surely dismissed workers who approached the court but did not stop at our information booth. Since the workers approaching the court had no way of know if the given day was assigned to treatment or control, we have no reason to believe there is differential non-compliance across treatment groups on either observable or unobservable characteristics. Hence, non-compliance in phase 3 is an issue of external rather than internal validity.[33]

[33] Even if every dismissed worker approaching the court had fully complied with the experimental protocol, the sample would not be representative of all aggrieved workers dismissed from Mexico City employers, because many dismissed workers obtain information from public or private lawyers without coming to the court. However, we believe the sample is relevant for understanding the effect of information provision on the decision to file case or settle.

Table A.2: Survey variables summary statistics

Panel A: worker characteristics			
	Mean (SD) N		Mean (SD) N
Age	44.37 (13.18) 102	Have you changed lawyer during trial?	0.09 (0.29) 102
Less than 50 employees	0.59 (0.49) 96	Probability of winning trial	79 (21.3) 102
Percentage of what is obtained	28.3 (11.11) 57	Most probable amount	97116.66 (355749) 102
Probability of winning	48.68 (34.06) 102	Most probable time	3.64 (9.76) 102
More than secondary education	0.58 (0.5) 102	How well were you treated?	0.45 (0.5) 102

Panel B: lawyer characteristics		
	Worker's lawyer	Employer's lawyer
	Mean (SD) N	Mean (SD) N
Age	37.79 (12.43) 231	36.56 (10.97) 223
Tenure	9.25 (8.88) 181	18.65 (143.32) 193
More than 100 historical cases	0.49 (0.5) 231	0.61 (0.49) 223
More than 30 current lawsuits	0.6 (0.49) 231	0.68 (0.47) 223
Less than 50 employees	0.65 (0.48) 173	0.46 (0.5) 151
Percentage of what is obtained	29.69 (7.79) 172	
Probability of winning trial	72.19 (20.14) 231	70.48 (21.53) 223
Most probable amount	237957.65 (1482119.93) 231	66884.79 (285657.2) 223
Most probable time	3.94 (2.3) 231	4.17 (2.54) 223

Notes: Summary statistics of survey variables.

Table A.3: Compliance Rate

(a) Phase 1

Group	N	Compliance with treatment				Compliance with baseline survey				Compliance with exit survey			
		Plaintiff	Defendant	Both	Any	Plaintiff	Defendant	Both	Any	Plaintiff	Defendant	Both	Any
Control	321	-	-	-	-	50.78	35.51	8.72	77.57	39.88	24.3	94.39	58.57
Calculator	315	77.65	73.03	75.81	79.72	48.25	33.65	12.38	69.52	38.63	25.86	88.47	54.83

(b) Phase 2

Group	N	Compliance with treatment				Compliance with survey			
		Plaintiff	Defendant	Both	Any	Plaintiff	Defendant	Both	Any
Control	329	-	-	-		-	-	-	-
Calculator	762	72.7	63.78	49.08	87.4	54.33	48.29	49.08	87.4

(c) % Shows up

	Phase 1		Phase 2	
	Control	Calculator	Control	Calculator
Employee	0.193	0.197	0.122	0.165
Emp Lawyer	0.847	0.832	0.85	0.853
Firm Lawyer	0.738	0.762	0.434	0.444

Notes: Compliance rate for each phase, both for treatment and survey. Each panel shows the percentage compliance with treatment and survey for plaintiff' side, defendant's side, both, and any. In phase 1, we did not measure compliance with treatment at the party level but rather at the casefile level, so that for example the 74.16 under the column "both" for compliance with treatment means that among the casefiles for which both parties showed up to the hearing, we were able to give the calculator to at least one of the parties in 74.16 percent of the cases. Also, in phase 1 we had an exit survey in addition to the baseline survey. In phase 2, we did measure compliance with the treatment by party. On the other hand, note from panel (b) that since in phase 2 control days were days that no implementing personnel were present in the subcourts, we have neither treatment nor survey compliance for the control group. Panel (c) shows from the court's administrative data who shows up to the hearings for each of the two phases, split by employee, employee lawyer, and firm lawyer.

Table A.4: Balance table

	Phase 1			Phase 2			Phase 1/2		
	C	T	p-value	C	T	p-value	C	T	p-value
Female	0.44	0.39	0.26	0.51	0.45	0.06*	0.47	0.43	0.09*
At will worker	0.18	0.18	0.93	0.1	0.07	0.2	0.14	0.1	0.04**
Weekly hours	56.58	58.05	0.28	54.97	56.64	0.05*	55.77	57.04	0.09*
Tenure at firm	104.05	236.84	0.04**	4.97	4.11	0.07*	56.96	69.32	0.57
Public Lawyer	0.1	0.12	0.3	0.06	0.06	0.76	0.08	0.08	0.85
% Reinstatement	0.45	0.44	0.83	0.48	0.51	0.29	0.46	0.49	0.24
% Severance pay	0.81	0.83	0.39	0.81	0.79	0.54	0.81	0.8	0.87
Daily wage	842.02	628.04	0.05*	617.35	600.52	0.85	728.57	608.02	0.08*
% Backpay	0.97	0.96	0.61	0.99	0.98	0.47	0.98	0.98	0.78
% Tenure bonus	0.78	0.83	0.08*	0.8	0.78	0.55	0.79	0.79	0.69
% Extra hours	0.68	0.71	0.43	0.71	0.77	0.04**	0.7	0.76	0.01**
% 20 days	0.31	0.32	0.72	0.35	0.34	0.7	0.33	0.33	0.86
% Sunday bonus	0.18	0.19	0.78	0.18	0.19	0.67	0.18	0.19	0.57
% Weekly rest	0.2	0.19	0.67	0.17	0.22	0.05*	0.19	0.21	0.18
%Mandatory rest	0.32	0.29	0.47	0.26	0.3	0.2	0.29	0.3	0.72
% Social security codef	0.53	0.53	0.98	0.63	0.64	0.73	0.58	0.61	0.23
% Earnings	0.36	0.35	0.86	0.37	0.35	0.54	0.36	0.35	0.57
% Nulity	0.62	0.57	0.23	0.61	0.57	0.22	0.61	0.57	0.08*
Entitlement	97065.2	71560.51	0.03**	63189.71	68890.8	0.5	78994.5	69581.42	0.18
Presence employee	0.19	0.19	0.94	0.16	0.19	0.14	0.17	0.19	0.31
Presence emp law	0.85	0.84	0.7	0.92	0.89	0.14	0.88	0.87	0.58
Presence firm	0.04	0.02	0.16	0.01	0.02	0.22	0.02	0.02	0.65
Presence firm law	0.74	0.77	0.41	0.79	0.78	0.76	0.77	0.78	0.54
Observations	330	326		327	768		657	1094	

Notes: Balance table on basic and strategic variables. The first three columns show phase 1, with means for each variable in the control, calculator, and conciliator treatment groups in the three columns. The next two columns show the control and treatment for phase 2, and finally the last two columns combine both phases. *, **, and *** indicates a difference that is significant at the 10%, 5%, and 1% level, respectively, as compared with the control group.

9.4 The Calculator

As described in the paper, one of the treatment arms required giving workers information about predicted outcomes for their case. In this subsection we describe the variables and machine-learning models we use to develop these predictions. Since the objective of the project was to provide the parties with the most accurate predictions possible, we considered several different models for each outcome of interest. The models were estimated using 70 percent of the data and then tested on the remaining 30 percent. We based the predictions on the model that yielded the lowest error in the test sample. Among the models we considered are the most common machine learning models, since these have shown in other settings to be very flexible and improve prediction accuracy.

We want to predict a series of outcomes, some of which are continuous and some of which are discrete. Different models are appropriate for these two types of outcomes, so we organize our discussion by first describing the discrete outcomes of interest and the models tested for those outcomes, and then describe the continuous outcomes of interest and the models we tested for those outcomes. We then describe the calculator templates used in both the first and second phase of the project.

For the calculators used in both phases, and for both the continuous and discrete models, we fed the models with the same set of input variables, all from the initial case filing. This is because, for operational reasons we could not use procedures that occur after the filing of the lawsuit, such as evidence submission. Also, the court wanted us to develop a parsimonious calculator that could be used in pre-judicial conciliation meetings prior to the plaintiff filing suit.

Discrete outcome models

The expected payment made to the plaintiff is a function of which party prevails and the amount transferred conditional on the outcome. There are five ways a case can end:

1. **Settlement:** The case may end with a voluntary agreement between the parties where the the plaintiff accepts a sum of money to cease the lawsuit and renounce the legal right to sue again for the same reason. To be valid, these settlements must be registered at the court, and therefore included in our administrative data.

2. **Court ruling with positive compensation:** Cases may proceed to a ruling by a judge that decides which side is wins the lawsuit and how much should be paid to the plaintiff. We classify an outcome as Court ruling with positive compensation if the case ends in a ruling by the judge and the worker actually collects a positive amount.[34]

3. **Court ruling with zero compensation:** The judge may also rule in the defendant's favor. In that event the worker receives nothing. However, the worker may also receive nothing if the judge rules in her favor, but she is unable to recover any of the judgment amount from the defendant. Defendants use a variety of strategies to avoid paying judgments, so the win but collect nothing outcome is not uncommon. The court records that we digitized did not allow us to differentiate between these two outcomes. We do not see this as an important shortcoming since from the point of view of the plaintiff what matters is the amount she receives.

[34] Because court rulings with positive collection are uncommon (3.3 percent of cases) we face the the problem of unbalancedness of our court rulings sample. To deal with this problem we used a Synthetic Minority Over-Sampling Technique (see Chawla et.al., 2002) and did a 80-20 train vs. test split on our data.

4. **Dropped:** A case can be dropped by the plaintiff at any time during the legal proceedings. The difference between dropped and settlement is that it can be done unilaterally by the plaintiff and no payment to the defendant is registered. Our understanding is that when cases are dropped it is because the plaintiff has little evidence to support the case. It is a decision of the plaintiff and not a mandate by the court.

5. **Expiry:** Finally, a labor suit may expire if the court requires information to continue the procedure, including items of evidence presented by one of the parties that the court needs to view before concluding the hearings procedure if proofs are not provided in a span of 4 months.

In our historical data of cases filed in 2011 and completed by the end of 2015, 63.3 percent end in settlement, 2.0 percent in a court ruling with positive collection, 6.6 percent with a court ruling with nothing collected, 20.3 percent expire and 7.6 percent are dropped.

We want to estimate the probability that a case with characteristics X_i ends in each of the five ways described above. We have a choice of estimating a single multinomial model or separate bivariate outcome models. We chose the bivariate option for simplicity. But to ensure that the probabilities summed to one, we set the probability of expiry equal one minus the sum of the probabilities of the other four outcomes.[35] We therefore estimated models for four bivariate outcomes, using each the following methods: (a) Logistic Regression, (b) Probit, (c) Random Forest[36], (d) Single-hidden-layer Neural Network (20 nodes in the hidden layer and 10% weight decay), (e) Gradient Boosting.[37] Models were estimated in a random sample training set made of 70 percent of the observation. The remaining 30 percent was used as a test set. For each model hyperparameters were chosen to minimize mean square error in the test set (MSE-T) using cross-validation. Once hyperparameters for each model were chosen, we chose among those optimized models bases on their mean square error. For each of the 5 discrete outcomes we want to predict we kept the model with the highest correlation between Y and $\hat{Y}$ in the test set.[38] Table A.5 shows the accuracy rate for each of these 5 outcomes and for each of the 5 models, highlighting in grey the one chosen. The correlations range from 0.61

[35] We also estimated a multinomial model. The results were similar, which is why we chose the bivariate models for simplicity.

[36] We performed grid search in order to find the best hyperparameter setting. We compared 7 different models with the number of trees ranging between 900 and 1500. Our final model resulted in a Random Forest of 1200 CARTs, which yielded an 86 percent accuracy rate on test classification.

[37] This was implemented with off the shelf models in R : e1071, randomForest, neuralnet, caTools, mboost.

[38] For the outcome of settlement, we find that the correlation between Y and $\hat{Y}$ is very slightly higher for Random Forest than for Logit, but we use the Logit model because it was simpler to implement in the field.

to 0.93 for the models selected.

Table A.5: Fit assessment of discrete calculator models

Phase 1: Discrete outcomes

	N	Logit	Probit	Random Forest	Neural Network	Gradient Boosting
Settlement	2075	0.61	0.61	0.62	0.57	0.40
Losing court ruling	2075	0.67	0.69	0.74	0.65	0.62
Winning court ruling	2075	0.67	0.69	0.75	0.70	0.67
Expiry	2075	0.93	0.93	0.93	0.93	0.93
Drop	2075	0.78	0.78	0.78	0.74	0.78

Phases 2 and 3: Discrete outcomes

	N	Logit	Probit	Random Forest	Neural Network	Gradient Boosting
Accuracy	432	0.71	0.72	0.89	0.80	0.75

Notes: Some statistics on the predictive power of the models considered for both -Phase 1 and Phase 2- calculator calibration processes. Statistics for models chosen for each problem are shown in bold. We show $Corr(y, \hat{y})$ for continuous outcomes. We considered random forests both for continuous and categorical outcome variables, using the algorithm's regression and classification methods, respectively.

Continuous outcome variables

Several relevant outcomes are continuous variables. We focus on three continuous variables to be provided to the parties:

1. *Amount collected conditional on a positive payment:* This is the peso amount actually collected by the plaintiff in the each of the two outcomes where they payments are positive: settlement and judgment in favor of the plaintiff, with recovery. The other endings all result in zero payments.

2. *Probability of Positive Recovery:* From the historical data for each casefile we know if the worker in fact was paid a positive amount or not (i.e. won the case in the court ruling and could indeed collect, or settled and got a positive amount).

3. *Duration of the case:* The number of months from the filing of the case to the date when it ended. As a significant share of the cases were not resolved by the end of 2015, we discuss censoring below.

For the continuous outcomes, we estimated a set of four different models for each of the two prediction problems.[39] The four models were: (a) OLS regression, (b) GLM Boosting, (c)

[39] Considering both regular and logarithmic models that would be eight different models for each outcome. The logarithmic models help to tackle the skewness of our dependent variable.

Random Forest[40], (d) Ridge Regression[41] [42]. As with the discrete variables, for each model the hyperparameters were chosen to minimize mean square error in the test set (MSE-T) using cross-validation. Once hyperparameters for each model were chosen, we chose among those optimized models based on their Mean Absolute Percentage Deviation (MAPD).

Finally we also wanted to predict the duration of the trial from initial suit to termination. We used the data with models similar to those described for the amount collected. However, for phase 2 the predictions from these models were very noisy: the models could not beat a simple average within each type of ending, so we decided to present the simple average within each type of termination rather than a model as a function of observables.

As with the discrete outcomes, Table A.6 shows the accuracy rate for settlement and duration outcomes and for each of the six models, highlighting in grey the ones chosen. The correlations range from 0.61 to 0.76 for the models selected for total compensation, but from 0.08 to 0.93 for the five duration outcomes.

[40] We perfomed grid search in order to find the best hyperparameter setting. We compared 7 different models with the number of trees ranging between 900 and 1500. Our final model resulted in a Random Forest of 1200 CARTs, which yielded an 86% accuracy rate on test classification.

[41] This was done with the libraries in R: e1071, randomForest, neuralnet, caTools, mboost.

[42] We also run the OLS and Random Forest models using logged data. In phase 2, we changed the models we estimated to include Kernel regression, running both OLS and Kernel in levels and logs. We do not present here the Boosted, Random Forest and Ridge regressions. The $corr(\hat{y^j}, \hat{y^k})$ between the predicted values of model j and model k for the union of all models is above 0.8 for all models.

Table A.6: Fit assessment of continuous calculator models

Phase 1: Continuous outcomes

Ending type	N	Regression	Log-regression	Boosted regression	Random Forest	Log-Random Forest	Ridge Regression
Total Compensation							
Settlement	1236	0.61	0.55	0.61	0.63	0.61	0.61
Court Ruling	66	0.70	0.76	0.70	0.49	0.28	0.67
Duration							
Settlement	1236	0.07	0.07	0.08	0.05	0.07	0.07
Losing court ruling	49	0.93	0.94	0.93	0.87	0.81	0.93
Winning court ruling	66	0.64	0.65	0.64	0.47	0.44	0.65
Expiry	118	0.35	0.34	0.37	0.29	0.39	0.42
Drop	468	0.20	0.16	0.19	0.13	0.17	0.05

Phases 2 and 3: Continuous outcomes

Ending type	N	Regression	Kernel regression	Log-regression	Log-kernel regression
Total Compensation					
Settlement	3130	0.59	0.46	0.69	0.63
Court ruling	105	0.44	0.30	0.18	-0.06

Notes: Some statistics on the predictive power of the models considered for both -Phase 1 and Phase 2- calculator calibration processes. Statistics for models chosen for each problem are shown in bold. We show $Corr(y, \hat{y})$ for continuous outcomes. We considered random forests both for continuous and categorical outcome variables, using the algorithm's regression and classification methods, respectively.

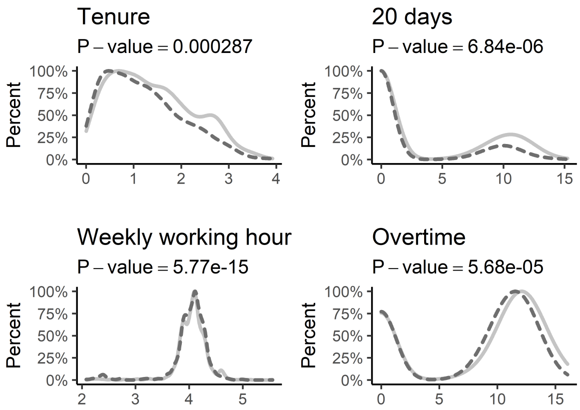
Other issues: Censoring

All of the models are estimated on a sample of cases filed in 2011 and completed by the end of 2015. The fact that the sample contains only the cases that were resolved could introduce a bias in the prediction of outcomes of ongoing cases. For example, if we are interested in the probability of winning *eventually* and if cases with larger expected payouts take longer to resolve, then excluding the unresolved cases may produce an underestimate of the average payments in all cases.

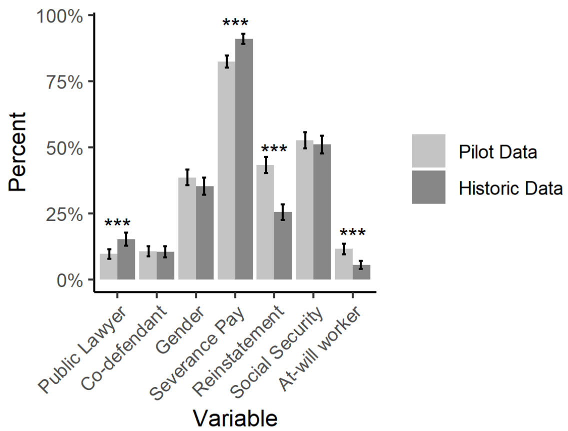
We are aware of this bias and it was communicated to the parties when the calculator information was provided. Although we cannot know how large the bias is, we performed two exercises that suggest it is not large. First, we compare characteristics of ongoing cases with those of the historical cases used in the models. In Figure A.1 we show that the two sets of cases are similar. The second exercise compares the characteristics of completed and continuing lawsuits within the historical data. To do this we drew a random sample of 956 cases filed in 2011 that were not finished by 2015 (i.e. this represents the complement of our historical dataset). We compare these 956 cases to the completed cases used to develop the models. Figure A.2 reports the results. There are no significant differences between these two samples, and the magnitude of the measured differences are all very small.

Figure A.1: Covariate distribution comparison : Historical and Phase 1 data

Continuous variables



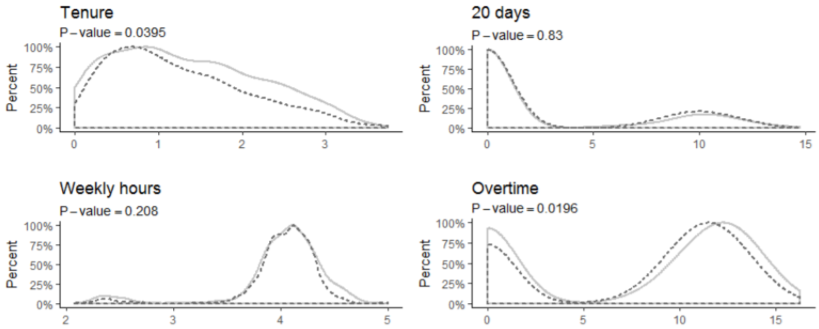
Categorical variables



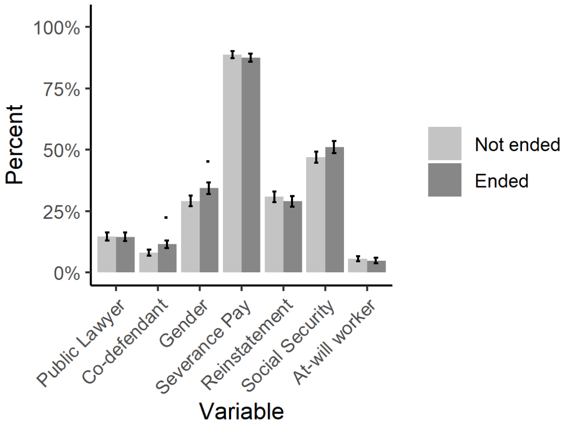
Notes: Covariate distributions comparisons. We compare between -used to calibrate our calculator- and Phase 1 data. All continuous variables are plotted in logs. Color guide is the same for both variable types. Plots for continuous variables show the p-value of a KolmogorovSmirnov test in the subtitle. In the categorical covariates plot, we show significance of a two-sided t-test of differences in means between samples.

Figure A.2: Covariate distribution comparison : Historical data: Ended and not ended cases

Continuous variables



Categorical variables



Notes: Covariate distributions comparisons between historical data cases: ended and not ended after 3 years. All continuous variables are plotted in logs. Color guide is the same for both variable types. Plots for continuous variables show the p-value of a KolmogorovSmirnov test in the subtitle. In the categorical covariates plot, we show significance of a two-sided t-test of differences in means between samples.

Appendix B

The Effect of Agency on Settlement Rates

We aim to write down a simple model that aims to illuminate why agency issues plausibly the underlying cause of the patterns in treatment we find on Tables 3 and 4. The treatment is effective only when plaintiffs are present to receive it directly. Moreover, this pattern is driven by cases where the plaintiff is represented by a private lawyer. These basic facts could be explained by several models, and we do not claim that the simple and highly stylized framework we write down here is the only way to rationalize them.

Private lawyers almost always receive a share of whatever the plaintiff collects. We should expect the contingency payment contract to align incentives with regard to effort at least partially. We focus instead on differences in incentives over the decision to settle or proceed with the case. The parties differ in their ability to diversify risks and are also likely to have different discount rates. Our framework focuses on these differences, and shows why they may lead to differences in preferences over taking a certain payoff now (i.e., a settlement) or an uncertain payoff later (i.e., a court judgment).

We start with three utility functions, $(U_w(x), U_l(x), U_f(x))$, defined over the amount awarded to plaintiffs for the plaintiff (w), the plaintiff's lawyer (l), and the firms(f)[43], and subjective distributions $g_k(x)$ for $k = \{w, l, f\}$ for the amount that party k recovers (pays) in a trial. We define the *settlement range* between the plaintiff and defendant as $I_{w,f} := \{x \mid \mathbb{E}_w U_w \leq x \leq \mathbb{E}_f U_f\}$, where expectations are taken with respect to $g_k(x)$ and the utility functions incorporate attitudes toward risk over uncertain outcomes. If the utility over outcomes differs for plaintiffs and their lawyers, then we can also define a similar settlement range between firms and plaintiff's lawyers as $I_{l,f} := \{x \mid \mathbb{E}_l U_l \leq x \leq \mathbb{E}_f U_f\}$.

We show three claims that follow from this simple framework.

Claim 1 *Suppose* $g_w(x) = g_l(x) = g_f(x) := g(x)$ *and*

$$A_{U_w}(x) \geq A_{U_l}(x) \geq A_{U_f}(x) \quad ; \forall\, x \in \operatorname{supp} g$$

where $A_U(x) - \frac{U''(x)}{U'(x)}$ *is the Arrow-Pratt measure of risk aversion. Suppose further that* $U'_w(x) \leq U'_l(x) \leq U'_f(x)$ *in a neighbourhood of* 0. *Then* $\emptyset \neq I_{l,f} \subseteq I_{w,f}$. *That is, the worker is more*

[43] We will assume that $U(0) = 0$, $U'(x) > 0$ and $U''(x) \leq 0$, so that U is risk averse or risk neutral. We can think of the 30 percent of any recovery that the plaintiffs' lawyers typically receive as being already reflected in their utility function.

willing to settle than her lawyer.

Proof of claim (1): It will suffice to prove that $\mathbb{E}_w U_w \leq \mathbb{E}_l U_l \leq \mathbb{E}_f U_f$.

Now,

$$\begin{aligned}
A_{U_w}(x) \geq A_{U_l}(x) &\iff \frac{U_w''(x)}{U_w'(x)} \leq \frac{U_l''(x)}{U_l'(x)} \\
&\iff \frac{d}{dx}\ln(U_w'(x)) \leq \frac{d}{dx}\ln(U_l'(x)) \\
&\iff \int_0^y \frac{d}{dx}\ln(U_w'(x))dx \leq \int_0^y \frac{d}{dx}\ln(U_l'(x))dx \ \ \forall\, y \in \ \operatorname{supp} g \\
&\iff \ln(\frac{U_w'(y)}{U_w'(0)}) \leq \ln(\frac{U_l'(y)}{U_l'(0)}) \ \ \forall\, y \in \ \operatorname{supp} g \\
&\iff \frac{U_w'(y)}{U_w'(0)} \leq \frac{U_l'(y)}{U_l'(0)} \ \ \forall\, y \in \ \operatorname{supp} g \\
&\implies U_w'(y) \leq U_l'(y) \ \ \forall\, y \in \ \operatorname{supp} g
\end{aligned}$$

As $U(0) = V(0)$ the last inequality tells us that

$$U_w(x) \leq U_l(x) \ \ \forall\, x \in \ \operatorname{supp} g$$

which together with the fact that all parties follow the same subjective distribution leads to

$$\mathbb{E}_w U_w \leq \mathbb{E}_l U_l$$

The other inequality is proved analogously.

Claim 1 says that even with the same subjective $g(x)$ for all parties, if plaintiffs are more risk averse than defendants[44] then the settlement range between the worker and firm is non empty. Moreover, if the worker is more risk averse than her lawyer, then the settlement range between the firm and the plaintiff's lawyer is a subset of the settlement range between the firm and the plaintiff. Therefore, the worker will be willing to accept some settlement offers that her lawyer will want to reject. The difference in the circumstances governing risk creates a potential conflict of interest between the plaintiff and her own lawyer.

[44] Since the workers typically have much lower incomes and have recently lost a job, and since both the plaintiff's lawyers and the firms typically manage several cases simultaneously, we believe this is a reasonable assumption. With a slightly more complex model, we can get the same result even with the same utility function for the lawyer and the worker, if lawyers are able to diversify across multiple cases. Since, on average, lawyers surveyed in phase 1 reported having more than 100 open cases, we think it is safe to say that lawyers are far more diversified than workers.

Claim 2 *Assume additionally that* $g_w^o \succeq_{FOSD} g_w$*. Then* $I_{w^o,f} \subseteq I_{w,f}$*. That is, overconfidence shrinks the settlement range.*

Proof of claim (2): As g_w^o first-order stochastically dominates g_w then

$$\mathbb{E}_{g_w^o} U \geq \mathbb{E}_{g_w} U$$

for all weakly increasing U. From this the claim follows directly.

Claim 2 relaxes the assumption of the same expectations over outcomes to allow for overconfidence. We show that if the worker is overconfident (i.e. g_w^o first-order stochastically dominates g_w) then the settlement range shrinks. Indeed, if the parties are sufficiently optimistic, the settlement range will be empty. Our surveys in phase 3 indicate that workers are wildly optimistic before filing a case. Because most workers are suing for the first time, while their lawyers have typically handled a large number of cases, the lawyers have the ability to guide workers toward realistic expectations or maintain their initial overconfidence. The misalignment between the worker and her lawyer generated by Claim 1 gives an incentive for the lawyer to inflate the expectations of the worker.

Claim 3 *Let the worker have a prior distribution* g_w*, and let the calculator be a signal with distribution* S*. Suppose further that the signal satisfies* $g_w \succeq_{FOSD} S$ *and that the agent is Bayesian. If we denote the posterior by* $\hat{g}_w$*, then* $I_{w,f} \subseteq I_{\hat{w},f}$*. That is, after updating the settlement range increases.*

Proof of claim (3): We will prove that $g_w \succeq_{FOSD} \hat{g}_w \succeq_{FOSD} S$, combining this fact with Claim 2 will yield the desired result.

As the agent is Bayesian, then by Theorem 1 in *Foundations for Bayesian Updating* we can write $\hat{g}_w = \alpha g_w + (1-\alpha)S$ for some $\alpha \in [0,1]$. Then,

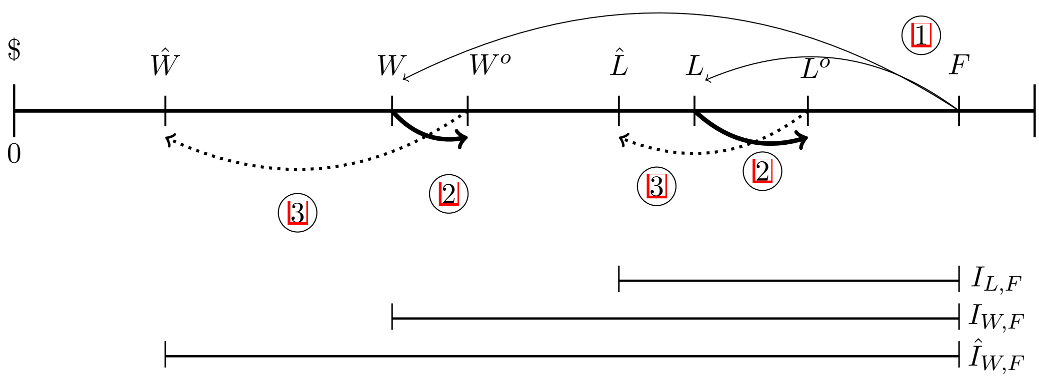
$$\mathbb{E}_{\hat{g}_w} U = (1-\alpha)\mathbb{E}_{g_w} U + \alpha \mathbb{E}_S U \leq (1-\alpha)\mathbb{E}_S U + \alpha \mathbb{E}_S U = \mathbb{E}_S U$$

for all weakly increasing U. As S is dominated by g_w, this yields the desired inequality.

Claim 3 introduces the calculator as a signal received by Bayesian updaters, and shows that when initially overconfident workers get the calculator, their settlement range increases to $\hat{I}_{w,f}$, leading to more settlements.

The effect of the three claims can be illustrated in the following figure, which shows the effect of overconfidence and differences in the ability to diversify risk on the settlement range over which parties might bargain a resolution to the case.

Figure B.1: Illustration of claims



Notes: To simplify the notation we will simply write the subindex $i \in \{W, L, F\}$ to denote $\mathbb{E}_i U_i$. The arrows describe the conclusion on each of the claims: Initially with the same subjective probability but with different degrees of risk aversion, the expected utility of the different parties differ. This is Claim 1. Claim 2 tells us that for overconfident individuals (here denoted by W^o), their expected utility increases. Finally, Claim 3 is about the updating process for lawyers and workers, assuming that the calculator reduces the overconfidence. The conclusion is that lawyers are *weakly* more overconfident than workers, so that they update less than workers. The last intervals denote the *settlement range*, the first one for the lawyer and the last two before and after updating, when the employee is present.

Appendix C

9.5 Tables

Table C.1: Amount asked (log), amount won (log), and probability of winning - Historic Data

Panel A : All cases				
	Total asked	Amount Won	Won/asked	Prob winning
Public Lawyer	-0.62***	0.0011	0.055**	1.65
	(0.033)	(0.22)	(0.024)	(2.33)
Constant	2.76***	5.48***	1.06***	75.6***
	(0.19)	(0.97)	(0.10)	(9.67)
Observations	4866	4866	4866	4866
BVC	YES	YES	YES	YES
Dummy Industry	YES	YES	YES	YES
R-squared	0.74	0.036	0.043	0.020
DepVarMean	11.7	6.40	0.21	65.3

Panel B : Settlement				
	Total asked	Amount Won	Won/asked	Prob winning
Public Lawyer	-0.67***	-0.20***	0.087***	0.018
	(0.041)	(0.053)	(0.020)	(0.37)
Constant	2.63***	6.59***	1.52***	99.4***
	(0.25)	(0.31)	(0.11)	(2.15)
Observations	3077	3077	3077	3077
BVC	YES	YES	YES	YES
Dummy Industry	YES	YES	YES	YES
R-squared	0.74	0.37	0.11	0.0068
DepVarMean	11.7	9.70	0.27	99.6

Panel C: Court ruling				
	Total asked	Amount Won	Won/asked	Prob winning
Public Lawyer	-0.44***	-0.22	0.027	-0.59
	(0.12)	(0.81)	(0.29)	(7.59)
Constant	2.14***	0.63	1.86**	7.23
	(0.68)	(3.14)	(0.84)	(27.5)
Observations	422	422	422	422
BVC	YES	YES	YES	YES
Dummy Industry	YES	YES	YES	YES
R-squared	0.78	0.13	0.065	0.13
DepVarMean	11.7	2.74	0.37	24.2

This table shows OLS regressions of log total amount asked in the initial labor suit, the amount actually won, the ratio of these two, and the probability of the worker recovering a positive amount. Panel A includes all our historical data files (i.e. for all types of case ending), Panel B focuses on cases ending in settlement, and Panel C on those ending in court ruling. All regressions control for our basic variable controls (gender, at will worker, tenure at the firm, daily wage, weekly hours worked), as well as industry dummies in which firm operates in. We have 4866 observations instead of 5005 since there are some missing values in the controls.

Table C.2: Balance of casefiles having negative recovery amount.

	Gender	At will worker	Tenure	Daily wage	Weekly hours
	(1)	(2)	(3)	(4)	(5)
Negative NPV	-0.035**	-0.034	-0.006***	-0.000	0.002***
	(0.014)	(0.028)	(0.001)	(0.000)	(0.001)
Constant	0.417***	0.404***	0.428***	0.402***	0.279***
	(0.017)	(0.016)	(0.017)	(0.016)	(0.035)
Observations	4503	4492	4470	4484	4425
R-squared	0.014	0.013	0.017	0.013	0.017
F stat	13.022	12.109	15.139	11.902	15.373
Court dummies	YES	YES	YES	YES	YES

Notes: Each column regresses a characteristic of the case against a dummy variable indicating that the plaintiff recovered an amount less than the MXN$2000 initial filing fee. The sample is all cases from the historical data using private lawyers.

Table C.3: Balance regression on characteristics conditional on employee present

	Control	Calculator	Observations	R-squared	p-value (calculator = 0)
Public Lawyer	0.223***	0.034	312	0.042	0.492
	(0.043)	(0.049)			
Gender	0.162*	-0.256**	312	0.080	0.0395
	(0.0890)	(0.121)			
At will worker	0.0197	-0.0351	312	0.049	0.742
	(0.0965)	(0.106)			
Tenure	0.189	-0.286*	306	0.068	0.0814
	(0.133)	(0.161)	.		
Daily wage	0.0659	-0.0975	307	0.006	0.493
	(0.141)	(0.141)			
Weekly hours	0.0230	-0.0203	306	0.136	0.863
	(0.0755)	(0.117)			

Notes: The table shows results of regressions with the specification $y_i = \alpha_t + \sum_{j=1,2} \beta_j T_j + \epsilon_i$, where j refers to the treatment assignment $calculator, conciliator$. The sample is limited to the cases in which the plaintiff was present at the hearing. All right-hand side variables are standardized so that the constant reflects balance in the control group. Each row represents a different regression with the indicated independent variable. The regressions all include subcourt fixed effects.

Table C.4: Employee presence

	Employee Present		
	Phase 1	Phase 2	Phase 1/2
	(1)	(2)	(3)
Female	-0.066*	0.000	-0.023
	(0.035)	(0.023)	(0.019)
Public Lawyer	0.317***	0.508***	0.416***
	(0.066)	(0.059)	(0.045)
At will worker	-0.063	0.013	-0.016
	(0.049)	(0.042)	(0.032)
Tenure (years)	0.004	0.005**	0.005***
	(0.003)	(0.002)	(0.002)
Daily wage	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Weeky hours	-0.002**	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
Age		-0.001	
		(0.003)	
Constant	0.313***	0.303***	0.298***
	(0.062)	(0.066)	(0.045)
Observations	520	1026	1546
R-squared	0.090	0.124	0.105
DepVarMean	0.208	0.183	0.191
p-value	0.000	0.000	0.000
p-value w/o PL	0.014	0.001	0.000

Notes: OLS of employee presence against basic variable controls as predictors. Last two rows shows the p-value testing the quality of all basic variables (& without public lawyer) respectively.

Table C.5: Expectations Relative to Prediction

Panel A				
	Expectation		Relative OC	
	Probability	Amount	Probability	Amount
	(1)	(2)	(3)	(4)
Employee's Lawyer (β_1)	0.070	11410.9	0.28	-0.17
	(0.044)	(13314.9)	(0.18)	(0.71)
Firm's Lawyer (β_2)	-0.25***	-21804.2*	-0.86***	-0.51
	(0.042)	(12614.9)	(0.17)	(0.67)
Constant (employee α)	0.74***	66755.3***	1.70***	1.00*
	(0.036)	(10791.5)	(0.15)	(0.58)
Observations	1274	1104	1177	1032
File Fixed Effects	YES	YES	YES	YES
R-squared	0.79	0.91	0.96	0.89
p-value:Emp Law	0	0	0	0
p-value:Firm Law	0	0	0	0.04

Panel B				
	Expectation		Relative OC	
	(1)	(2)	(3)	(4)
	Probability	Amount	Amount	Amount
Mean	0.885***	58,077***	1.713***	1.315***
	(0.00288)	(1,794)	(0.0729)	(0.0619)
Standar deviation	0.166	74471	3.026	2.570
Observations	3,330	1,723	1,723	1,723
Ignores prob. (%)	13.08			
Standard deviation	0.166	74471	3.026	2.570
Reference			Mid prediction	Max prediction

Notes: The table regresses measures of expectation elicited in the baseline survey on dummies of who is the respondent of the survey. For some cases we could elicit the expectation of more than one party (employee, employee's lawyer, firm's lawyer). The omitted variable is the employee dummy, so the interpretation of the employee's lawyer and firm's lawyer coefficients are relative to the employee who is captured in the constant. It combines two phases in one singled pooled dataset. Each column represents a different regression. The first column use elicited probability of winning as a dependent variable. The exact question is: *"How likely is it that you will win the lawsuit if it ends in a court judgment?"*. The second column use the peso amount (undiscounted) that they expect to recover conditional on winning. The exact wording of the survey question is: *"in case you win the lawsuit, what amount are you most likely to win?"*. All columns include casefile fixed effects, avoiding the comparison across casefiles. The bottom of the table present the p-values of two null hypothesis: $\alpha + \beta_1 = 0$, and $\alpha + \beta_2 = 0$, telling us whether the employee's laywer, or firm's lawyer are more or less confident. Last 2 columns use a measure of "overconfidence", which compares the subjective expectation vs the personalized calculator prediction. Relative OC is computed as $\frac{expectation-prediction}{prediction}$, where expectation refers to the expectation measured in the baseline survey.

Table C.6: First stage and robustness for the control function regression

	Probit	First Stage (OLS)	First Stage (probit)		CF
	(1)	(2)	(3)	(4)	(5)
Calculator	0.122	0.028	0.0281	0.110	0.002
	(0.091)	(0.017)	(0.0183)	(0.069)	(0.015)
Emp present (EP)	0.726***				0.698***
	(0.174)				(0.199)
Calculator#EP	0.253				0.122**
	(0.202)				(0.059)
Control Function					-0.312***
					(0.104)
Time instrument		0.055***	0.0550***		
		(0.018)	(0.0178)		
09:30				-0.133	
				(0.159)	
10:00				-0.327***	
				(0.113)	
10:30				-0.157	
				(0.134)	
11:00				-0.224**	
				(0.105)	
11:30				-0.504***	
				(0.184)	
12:00				-0.139	
				(0.144)	
12:30				-0.031	
				(0.441)	
Constant	-1.969***	-0.072	-0.0721	-2.134***	0.067
	(0.510)	(0.049)	(0.0500)	(0.464)	(0.061)
Observations	1,709	1,732	1,732	1,732	1,732
R-squared	.	0.034	.	.	0.144
Casefile controls	YES	YES	YES	YES	YES

Notes: The first column repeats the regression in column 5 of 3 using a probit specification rather than a linear probability model. The second column is a OLS first stage for employee present included just for comparison with columns 3 and 4. The third column is the first stage of the control function regression shown in column 9 of 3. Columns 4 and 5 repeat the control function regression using individual dummy variables for each-hour hearing time. Column 4 shows the first stage and column 5 the outcome regression including the residual from the regression in column 4. See the notes to 3 for further details.

Table C.7: Heterogeneity in treatment effects

Interaction var	Daily wage			Tenure			Weekly hours		
	Phase 1	Phase 2	Phase 1/2	Phase 1	Phase 2	Phase 1/2	Phase 1	Phase 2	Phase 1/2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control (Constant)	0.054***	0.074	0.079**	0.055***	0.10**	0.074**	0.067***	0.100**	0.10***
	(0.018)	(0.045)	(0.036)	(0.019)	(0.040)	(0.035)	(0.023)	(0.044)	(0.033)
Calculator	0.096**	0.044	0.0013	0.089**	0.034	0.0068	0.068	0.027	0.0023
	(0.041)	(0.027)	(0.020)	(0.038)	(0.022)	(0.019)	(0.047)	(0.038)	(0.027)
Interaction Var (Int)	0.045	0.084	0.039	0.036	0.034	0.048*	0.019	0.019	-0.023
	(0.034)	(0.060)	(0.040)	(0.031)	(0.037)	(0.025)	(0.033)	(0.044)	(0.025)
Calculator#Int	-0.070	-0.00019	0.027	-0.065	0.0054	0.012	-0.0091	0.035	0.027
	(0.065)	(0.066)	(0.046)	(0.056)	(0.040)	(0.030)	(0.059)	(0.063)	(0.042)
Emp present (EP)			0.045			0.15***			-0.011
			(0.050)			(0.047)			(0.067)
Calculator#EP			0.32***			0.22***			0.30***
			(0.079)			(0.067)			(0.095)
EP#Int			0.14*			-0.0041			0.19**
			(0.077)			(0.072)			(0.091)
Calculator#EP#Int			-0.24**			-0.13			-0.16
			(0.10)			(0.093)			(0.13)
Observations	507	1067	1574	506	1041	1547	502	1074	1576
R-squared	0.015	0.060	0.13	0.012	0.053	0.14	0.011	0.054	0.13
DepVarMean	0.10	0.21		0.10	0.21		0.11	0.20	

Interaction var	Gender			Public Lawyer			Entitlement		
	Phase 1	Phase 2	Phase 1/2	Phase 1	Phase 2	Phase 1/2	Phase 1	Phase 2	Phase 1/2
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Control (Constant)	0.074***	0.092**	0.092***	0.068***	0.11***	0.095***	0.056***	0.054	0.064*
	(0.022)	(0.040)	(0.030)	(0.018)	(0.029)	(0.027)	(0.021)	(0.042)	(0.034)
Calculator	0.086**	0.056**	0.0017	0.060*	0.045*	0.015	0.069*	0.060*	0.0051
	(0.040)	(0.024)	(0.017)	(0.030)	(0.023)	(0.016)	(0.037)	(0.030)	(0.020)
Interaction Var (Int)	0.0033	0.049	0.013	0.065	0.041	-0.029	0.033	0.11**	0.066*
	(0.036)	(0.041)	(0.023)	(0.060)	(0.11)	(0.042)	(0.035)	(0.049)	(0.036)
Calculator#Int	-0.063	-0.025	0.029	-0.0059	-0.032	0.0015	-0.018	-0.022	0.019
	(0.062)	(0.039)	(0.027)	(0.083)	(0.12)	(0.036)	(0.057)	(0.063)	(0.043)
Emp present (EP)			0.095*			0.12**			0.093
			(0.055)			(0.046)			(0.057)
Calculator#EP			0.27***			0.23***			0.25***
			(0.064)			(0.069)			(0.089)
EP#Int			0.070			0.077			0.063
			(0.089)			(0.096)			(0.092)
Calculator#EP#Int			-0.22*			-0.24**			-0.12
			(0.11)			(0.11)			(0.13)
Observations	514	1081	1595	514	1081	1595	499	1081	1580
R-squared	0.014	0.051	0.13	0.014	0.050	0.13	0.012	0.064	0.14
DepVarMean	0.11	0.21		0.11	0.21		0.10	0.21	

Notes: We test for heterogeneity of treatment effects by interacting treatment with the variable shown in the column header. For each test, the first column uses the sample from phase 1, the second from phase 2, and the third from the combined sample. For example, the third column in the top panel shows that, in the combined sample, plaintiffs with higher daily wages (as reported in the case file) were marginally significantly less likely to settle when the plaintiff was present and provided the calculator information.

Table C.8: Treatment Effects with placebo arm - Phase 1

	Months after treatment	
	Same day	
	(1)	(2)
Control (Constant)	0.072***	0.041***
	(0.015)	(0.013)
Calculator	0.044*	0.016
	(0.024)	(0.020)
Placebo	-0.011	0.0058
	(0.020)	(0.018)
Placebo ctrl	-0.032*	-0.017
	(0.017)	(0.014)
Emp present (EP)		0.16***
		(0.053)
Calculator##EP		0.15*
		(0.086)
Plaecbo##EP		-0.050
		(0.077)
Placebo ctrl##EP		0.047
		(0.076)
Observations	1668	1668
R-squared	0.012	0.096
DepVarMean	0.064	0.064
InteractionVarMean		0.13
Calc=Placebo	0.017	0.62
Calc=Conc=Placebo=0	0.070	0.43
=Placebos	0.10	0.12

Notes: The table reports regressions using the full sample and including variables indicating that the case received a placebo treatment, as described in the text.

Table C.9: Treatment generated updating in probability - Phase 1

(a) Overconfidents

	Relative updating		
	Employee	Emp Lawyer	Firm Lawyer
	(1)	(2)	(3)
Calculator	-0.45**	0.63	-0.48
	(0.21)	(3.92)	(0.62)
Constant	-1.63**	-15.9	1.24*
	(0.58)	(11.1)	(0.72)
Observations	30	79	58
Basic Variable Controls	YES	YES	YES
Other controls	YES	YES	YES
R-squared	0.45	0.11	0.095
Update (mean)	-0.18	2.58	0.65
Update (SD)	0.46	16.2	2.07

(b) Underconfidents

	Relative updating		
	Employee	Emp Lawyer	Firm Lawyer
	(1)	(2)	(3)
Calculator	14.5	-2.59	0.042
	(15.6)	(1.97)	(0.36)
Constant	-167.5*	3.11	-2.46**
	(72.1)	(5.50)	(1.07)
Observations	20	56	32
Basic Variable Controls	YES	YES	YES
Other controls	YES	YES	YES
R-squared	0.97	0.15	0.16
Update (mean)	18.1	2.20	-0.23
Update (SD)	77.9	9.27	0.98

Notes: This table presents the treatment effect of the calculator on relative expectation updating, defined as $\frac{exitsurvey - initialsurvey}{initialsurvey}$. Sample is restricted to over-confident cases (i.e. *initial survey¿calculator prediction*) for Panel A, while it is restricted to under-confident cases for Panel B. For both panels, column 1 presents the results for surveyed employees, column 2 the results for surveyed employee's lawyer and table 3 the results for the firm's lawyer. All regressions include basic variables as controls (Public lawyer, Gender, At will worker, Tenure, Daily wage & Weekly hours).

Table C.10: Duration of Cases by Treatment

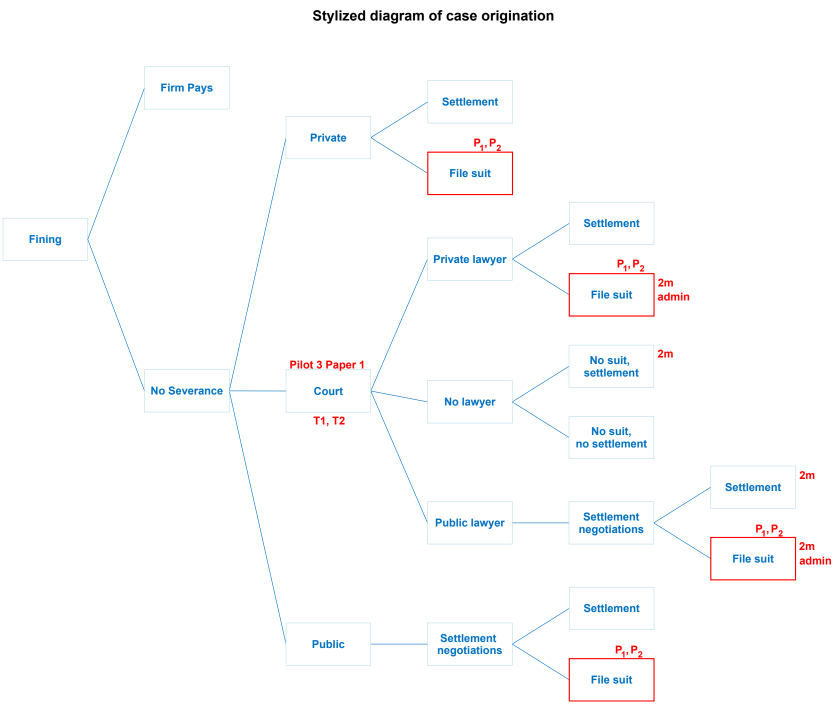
Case duration in months			
	OLS	COX	COX
	(1)	(2)	(3)
Calculator	17.480	-0.073	-0.041
	(26.680)	(0.065)	(0.101)
Emp present (EP)	-59.887	0.133	0.013
	(51.101)	(0.132)	(0.210)
Calcularot#EP	-111.731*	0.318*	-0.071
	(63.117)	(0.183)	(0.277)
Observations	1,549	1,549	816
R-squared	0.427		
DepVarMean	0.174		
Casefile Controls	Yes	Yes	Yes
Includes settled	Yes	Yes	No

Notes: The table shows the duration of cases in phases 1 and 2, measured in days from the date of filing to the date of the final resolution. All regressions control for the age of the case at the time of treatment, and include dummies for phase 2 and the subcourt where the case was heard. Column 1 is an OLS regression and Columns 2 and 3 are Cox proportional hazards models, showing the coefficients rather than proportional hazards. Columns 1 and 2 include all cases and Column 3 removes from the sample cases that were settled. The sample in all regressions is limited to cases with private lawyers. The sample is slightly smaller than that in Table 4 because we eliminate a few cases with a missing date of filing.

9.6 Figures

Figure C.1: Stylized Depiction of the Labor Justice Process

Diagram of Case Origination.pdf Diagram of Case Origination.pdf



Notes: This figure is a stylized representation of the process that casefiles follow in Mexico City's Labor Court.

Figure C.2: Calculator Treatment Format (example) - Phase 2

(a) Plaintiff

CÁLCULO DE COMPENSACIÓN TRABAJADOR

ALEJANDRA [redacted] vs [redacted] SA DE CV
Expediente J11 / [redacted] / 2016

- **La Calculadora se basa en 5000 expedientes concluidos del año 2011.**
- **A continuación le da información de cómo terminaron juicios similares al suyo.**
- **¡IMPORTANTE! 16% de los juicios duran más de 4 años.**

De los juicios concluidos más similares al suyo…	
→ Los que llegaron **a un convenio**, en promedio obtuvieron:	$22,494
→ De los que continuaron hasta un **laudo**, el **70%**	**NO** obtuvieron pago.

- **Como referencia, le damos los datos básicos de su demanda**

Datos básicos de la demanda		
Género: Mujer	**Salario diario:** $700.00 mxn diarios	**Antigüedad:** 1.85 años
Prestaciones Solicitadas	Reinstalación	✓
	Indemnización Constitucional (90 días de salari	✓
	Horas Extras	✓
	Indemnización de 20 dias por año	✓

La Calculadora provee datos estadísticos de juicios similares. No afecta sus derechos ni influye en su juicio.

No. Folio : J11 / [redacted] / 2016 A

(b) Defendant

CÁLCULO DE PAGO DEMANDADO

ALEJANDRA [redacted] vs [redacted] SA DE CV
Expediente J11 / [redacted] / 2016

- **La Calculadora se basa en 5000 expedientes concluidos del año 2011.**
- **A continuación le da información de cómo terminaron juicios similares al suyo.**
- **¡IMPORTANTE! 16% de los juicios duran más de 4 años.**

De los juicios concluidos más similares al suyo…	
→ En caso de **perder en el laudo**, en promedio pagan	$174,197
→ En caso de **llegar a un convenio**, en promedio pagan	$22,494

- **Como referencia, le damos el cálculo del 100% de las prestaciones básicas, en caso de que gane el trabajador por despido injustificado**

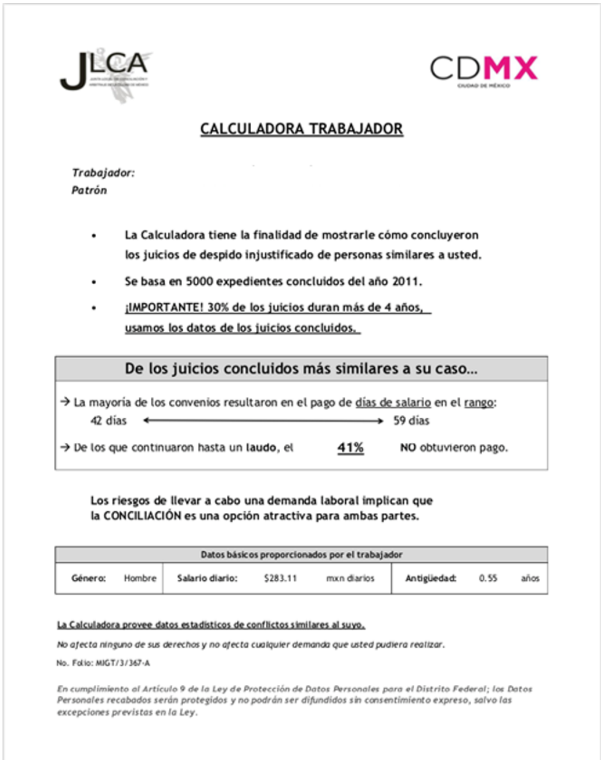
Datos del trabajador		
Género: Mujer	**Salario diario:** $700.00 mxn diarios	**Antigüedad:** 1.85 años
Indemnización Constitucional	90 días de salario diario integrado	$63,000.00
Prima de Antigüedad	12 días por año laborado	$3,246.58
Aguinaldo	Parte proporcional de 15 días del último año	$7,312.33
Vacaciones	Parte proporcional del último año laborado	$4,815.12
100% DE LAS PRESTACIONES BÁSICAS		**$78,374.03**

La Calculadora provee datos estadísticos de juicios similares. No afecta sus derechos ni influye en su juicio.

No. Folio : J11 / [redacted] / 2016 D

Notes: The figure shows examples of the calculator treatment used in phase 2. The left panel format was given to the plaintiff / dismissed worker and the right panel format to the defendant. The calculator shows two numbers: first, for cases with similar characteristics that settled, the average amount obtained in settlement; second, each party's contingency in case they did not settle and proceeded to a judge's ruling. This is the likelihood of not obtaining any payment for the plaintiff and the recovery amount conditional on positive recovery for the defendant. As in phase 1, parties were told that this information comes from a statistical exercise based on completed historical cases, and that it gives average prediction based on variables of the initial lawsuit described in the calculator treatment.

Figure C.3: Calculator Treatment Format (example) - Phase 3



CDMX

CALCULADORA TRABAJADOR

Trabajador:
Patrón

- La Calculadora tiene la finalidad de mostrarle cómo concluyeron los juicios de despido injustificado de personas similares a usted.
- Se basa en 5000 expedientes concluidos del año 2011.
- ¡IMPORTANTE! 30% de los juicios duran más de 4 años, usamos los datos de los juicios concluidos.

De los juicios concluidos más similares a su caso...

→ La mayoría de los convenios resultaron en el pago de días de salario en el rango:
42 días ⟷ 59 días

→ De los que continuaron hasta un **laudo**, el **41%** **NO** obtuvieron pago.

Los riesgos de llevar a cabo una demanda laboral implican que la CONCILIACIÓN es una opción atractiva para ambas partes.

Datos básicos proporcionados por el trabajador								
Género:	Hombre	Salario diario:	$283.11	mxn diarios	Antigüedad:	0.55	años	

La Calculadora provee datos estadísticos de conflictos similares al suyo.

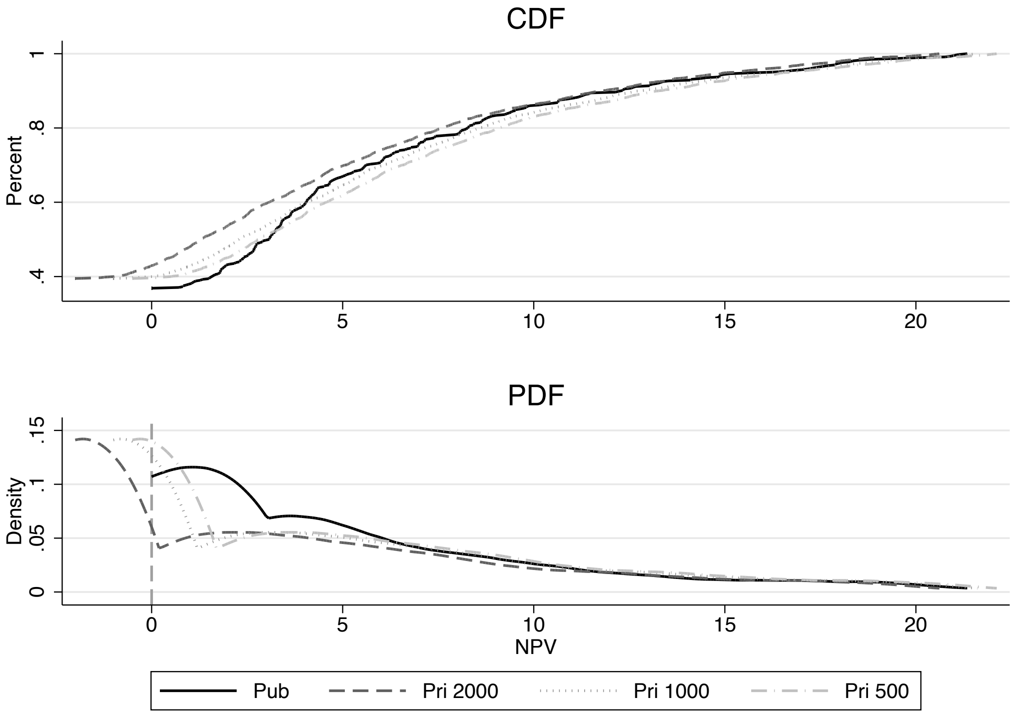
No afecta ninguno de sus derechos y no afecta cualquier demanda que usted pudiera realizar.

No. Folio: MIGT/3/367-A

En cumplimiento al Artículo 9 de la Ley de Protección de Datos Personales para el Distrito Federal; los Datos Personales recabados serán protegidos y no podrán ser difundidos sin consentimiento expreso, salvo las excepciones previstas en la Ley.

Notes: The figure shows an example of the calculator treatment provided to dismissed workers in phase 3. The calculator shows the inter-quartile range of settlement amounts expressed as the number of days paid and the percentage of cases continuing to court judgment where the plaintiff recovered nothing. The information provided is specific to the worker's characteristics. As in phases 1 and 2, parties were told that this information comes from a statistical exercise based on completed historical cases, and that it gives average prediction based on variables of the initial lawsuit described in the calculator treatment.

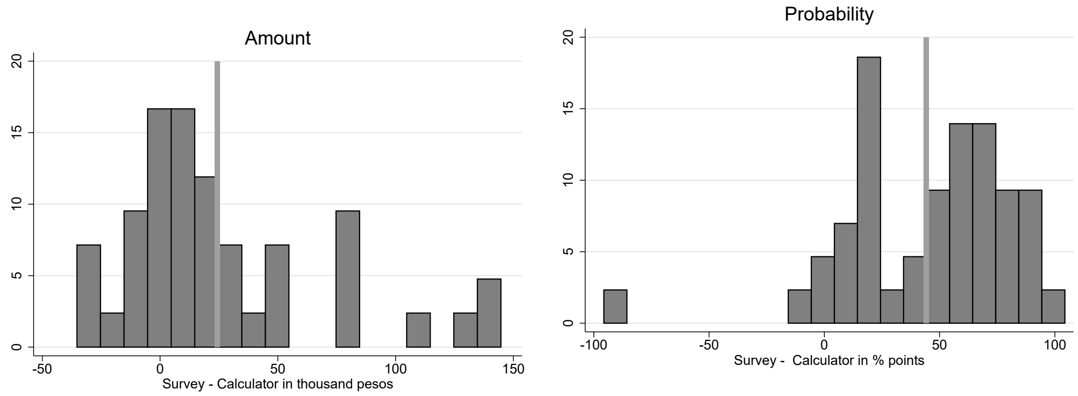
Figure C.4: Distribution of Amount Collected, by Type of Lawyer



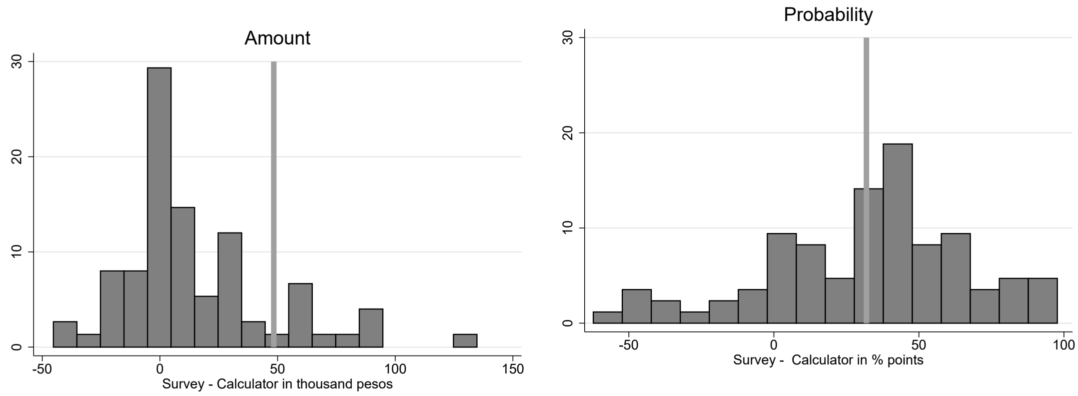
Notes: This Figure uses the historical data to show cumulative distributions and densities of the amount received in the historical data. It uses all casefiles endings (court ruling, settlements, drop and expiry). Amounts on the x-axis are in thousand pesos and brought to present value to the time of suing, with a monthly interest rate of 3.43. Since we care about what the worker actually receives, when they use a private lawyer we subtract a 30% of the recovery and the initial filing feel. The graph shows outcomes for three different initial fees (indicated in the legend): MXN$2,000 pesos, MXN$1,000 pesos, and MXN$500 pesos. The modal fee in the survey data is MXN$2,000 pesos. To the left of the vertical line at zero the worker loses money (from the initial fee). The figure indicates that there are a large fraction of cases where the worker has a negative net recovery.

Figure C.5: Subjective expectation minus prediction - Phase 1

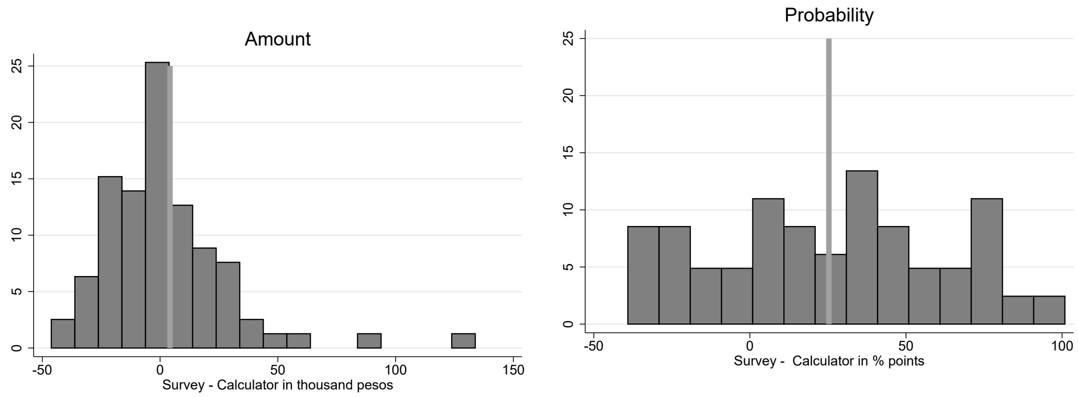
(a) Employee



(b) Employee's Lawyer

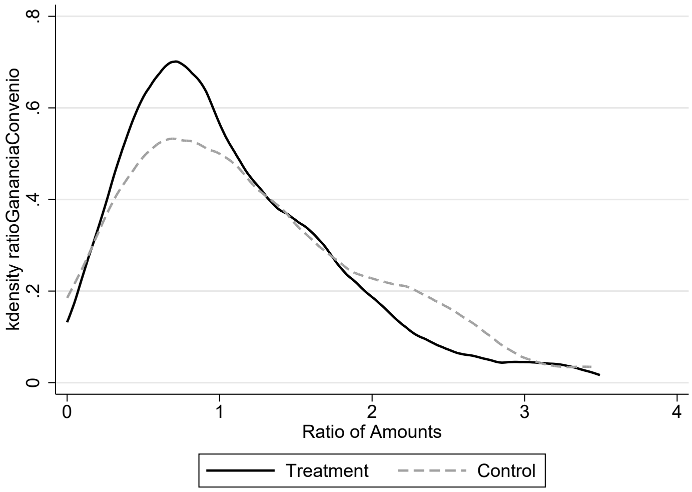


(c) Firm's Lawyer



Notes: Difference in thousand pesos for amounts (panel on left) and percentage points for probabilities (panel on right) from what the subject expects vs. what our models predict. Note how, for Employee and employee lawyer, the distribution for amounts is much more skewed to the right than for firm lawyers. This is only natural, since the former are thinking about expected wins and the latter about expected losses.

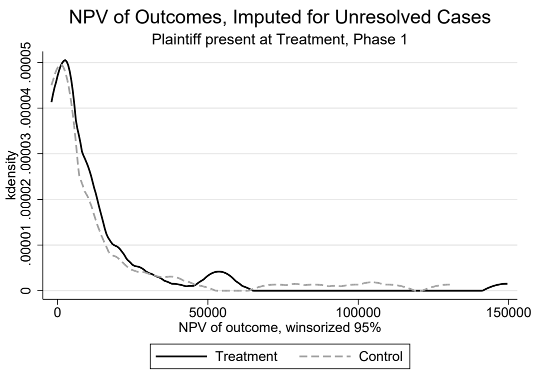
Figure C.6: Settlement Amount vs. Calculator



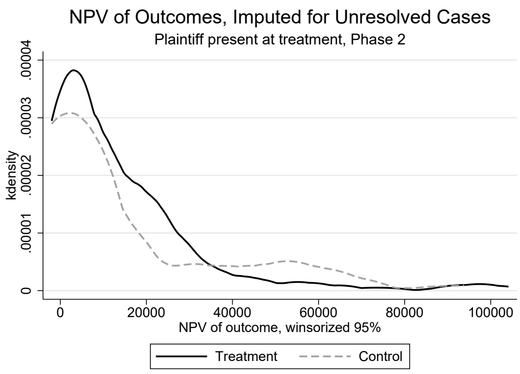
Notes: This figure presents the distribution of actual settlement amount over the calculator prediction. Data are truncated at the 99th percentile to compress the scale.

Figure C.7: Outcomes when Plaintiff was Present, by Treatment

(a) Phase 1



(b) Phase 2

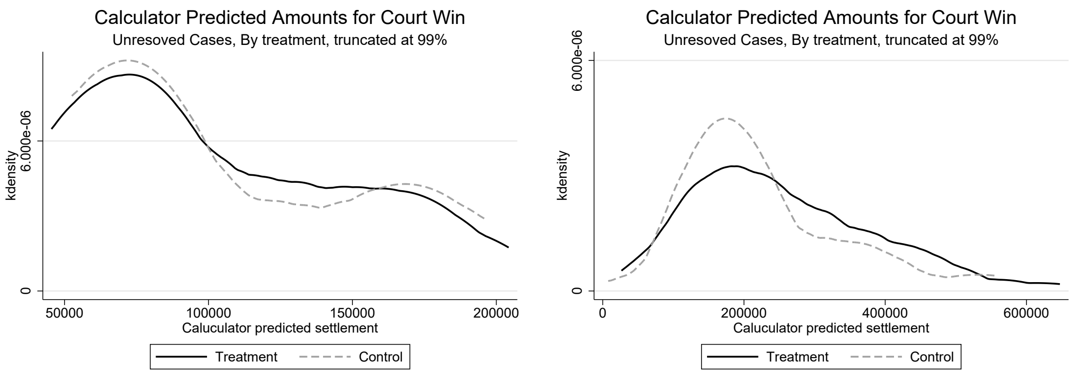


Notes: This figure shows the kernel density of the NPV of the amount awarded by treatment and phase of the intervention. NPV is discounted at an annual rate of 50 percent and the value of unresolved cases is imputed from the calculator prediction for outcome by court judgment. Data are winsorized at the 95th percentile.

Figure C.8: Calculator Predictions for Plaintiff Court Judgment, By Phase

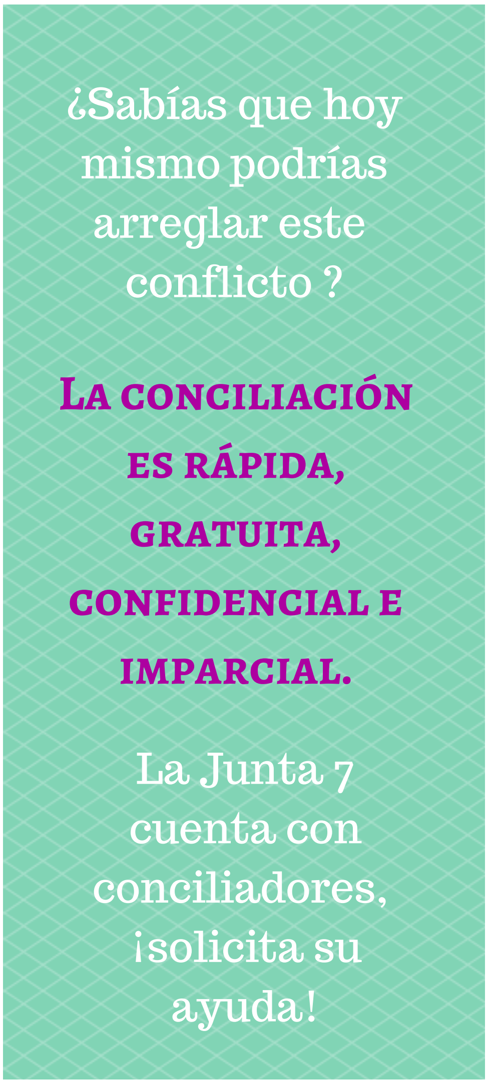
(a) Phase 1

(b) Phase 2



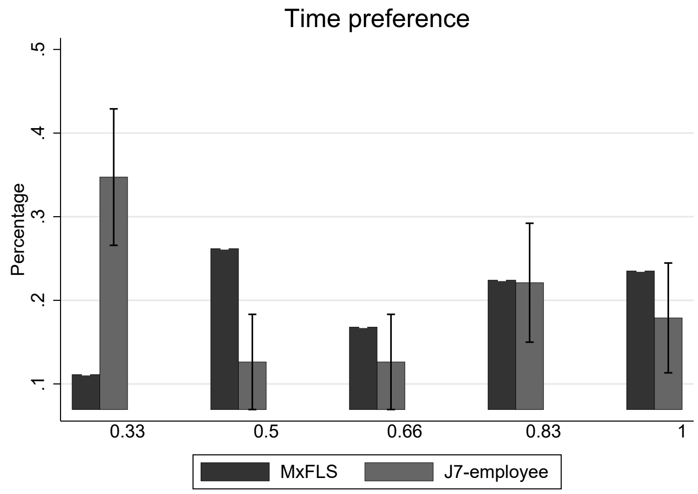
Notes: This figure presents the calculator prediction for amount recovered conditional on the plaintiff winning a court judgment, for treatment and control cases that were unresolved as of 2020. The cases in phase 1 and phase 2 are shown separately. Data are truncated at the 99th percentile to compress the scale.

Figure C.9: Information handed out in placebo test



Notes: This is the information brochure handed out to subjects in the placebo test. Essentially, it is a reminder of the existence of the conciliation process, which is free, confidential and unbiased.

Figure C.10: Discount rate for Phase 1 and MxFLS



Notes: Comparison of discount rates for Phase 1 data and survey data from the MxFLS (Mexican Family Life Survey- a longitudinal survey in Mexico that follows individuals across rounds). In both surveys individuals are asked whether they would choose to recieve \$1,000 dolars today or an array of different values in the future (\$1,200, \$1,500, \$2,000 or \$3,000 dollars). The monthly discount rate is calculated by dividing the \$1,000 dollars over the answered choice made by individuals. This was done accordingly to the MxFLS.