The Quality of Vote Tallies: Causes and Consequences*

Cristian Challú
ITAM

Enrique Seira
ITAM

Alberto Simpser
ITAM

February 2019

Preliminary Draft, please do not circulate or cite

Abstract

The credibility of election outcomes hinges on the accuracy of vote tallies. We provide causal evidence on the drivers and the downstream consequences of variation in the quality of vote tallies. Using data for the universe of polling stations in Mexico in five national elections, we document that over 40 percent of polling-station-level tallies display inconsistencies. Our evidence strongly suggests these inconsistencies are non-partisan. Using an original survey of 80,000 poll workers together with detailed administrative data, we find that inconsistencies cause recounts and lower trust in electoral institutions. Finally, using data for more than 1.5 million poll workers we show that lower educational attainment, higher workload, and higher complexity of the tally cause more inconsistencies. We discuss policy implications of our findings.

Keywords: Elections, vote counting, vote tallies, mistakes, recounts, democracy.

*Comments welcome. Challú: cristiani.challu@gmail.com, Seira: enrique.seira@itam.mx, Simpser (corresponding author): alberto.simpser@itam.mx.

1 Introduction

The credibility of election outcomes, and the health of democracy, hinge on the accuracy of vote tallies. Vote counting, however, is generally inaccurate. Whether inaccuracies are small or large, and whether they result from willful malfeasance or from unwitting error, they constitute political dynamite susceptible to exploitation for partisan ends. Disputes over the accuracy of the vote count in the 2000 USA presidential election, for example, had aftereffects that linger in the American political environment to date. In Ecuador's 2017 presidential elections, arithmetic and numerical inaccuracies in the vote tallies were used by the runner-up to push for a large-scale recount. And the call to recount "vote by vote, precinct by precinct" after the 2006 presidential elections in Mexico promoted long-lasting mistrust of the electoral system among a large fraction of the citizenry.[1]

Inaccuracies in the vote count, of course, could stem from fraudulent electoral practices (Myagkov et al., 2010; Hyde, 2007; Cantú, 2018). But even in a clean election, the imperfect nature of the counting process makes it impossible to guarantee the accuracy of the tally. Machine-based vote counts have been shown to be inaccurate (Alvarez et al., 2009), and the problem is likely graver when human error is potentially involved (Ansolabehere and Reeves, 2004; Goggin et al., 2012; Alvarez and Hall, 2008). Yet hand-counting is the rule in almost every country with elections (ACE Project). In fact, hand counting is making a comeback even in places where electronic voting used to be the rule, due to concerns about foreign meddling and hacking.[2] Nevertheless, we know little about the consequences of inaccurate tallies, and about the causes of such inaccuracies when votes are counted by people.

This paper presents what we believe to be the first systematic evidence on the prevalence, the consequences, and the causes of inaccuracies in the hand-counting of votes in mass elections. Our empirical analysis is based on a unique dataset covering the universe of polling stations, poll workers, and party representatives in five national elections in Mexico in 2009-2015. Altogether, we observe over 600 thousand polling-station-level tallies, over 1.6 million citizen poll workers, and hundreds of thousands of party representatives at the polls. Additionally, we conducted an original survey of citizen attitudes towards the electoral authorities on 80 thousand citizen poll workers.

Information on inaccuracies is culled from the official document that polling-station workers must fill out by hand, on paper, at the end of Election Day after counting the ballots in their corresponding polling station. This document, known as an *acta* (which we translate as *tally*), constitutes the basic input used by the electoral authorities to compute official

[1] On the 2000 USA election see Posner (2000). On Ecuador see Páez Benalcázar (2017), https://tinyurl.com/y8maj6yv. On Mexico see https://tinyurl.com/y9k9n4ww, https://tinyurl.com/ybuqvxeh, https://tinyurl.com/ycfyfj3x, https://tinyurl.com/ycl526lu.

[2] https://tinyurl.com/y9dwu98s and https://tinyurl.com/y8unqqhp

election results. Our measures of inaccuracies follow the electoral authority's own definitions of inconsistencies in the vote tallies. An inconsistency is said to exist when two or more fields in the tally that should satisfy an accounting equality fail to do so. In any given polling station, for example, the number of cast ballots plus the number of unused ballots should equal the initial number of ballots.

We find that inconsistencies in vote tallies are remarkably common, being present in more than two out of five *actas*, and in a similar proportion of polling stations.[3] We find no evidence, however, that tally inconsistencies in the period we study are the result of partisan malfeasance. This is consistent with the viewpoint of the Mexican electoral authorities, as well as with the scholarly consensus that Mexican elections have been virtually free of many traditional forms of election fraud since at least 1997 as a result of the major electoral reforms of the mid-1990s.[4]

Even honest mistakes in the tallying of votes, however, can have fateful consequences. As illustrated by the cases of the USA, Ecuador and Mexico mentioned previously, inaccuracies in vote tallies are often seized upon by losing parties in countries the world over to impugn the credibility of election results and, in some cases, that of electoral authorities or institutions themselves. In many countries, inconsistencies provide a legal basis for recount requests, as is the case in Argentina, Austria, Brazil, Chile, Colombia, Denmark, Ecuador, Honduras, Mexico, and Spain among others (ACE Project). Inconsistent tallies have led to major court cases in Armenia, Mali, Mexico, and the USA, to name a few examples (Posner, 2000; Autheman, 2004). Our findings are consistent with these observations: in the elections we study, inaccurate tallies are associated with a 22pp-greater probability that the relevant polling station's votes are recounted. We also find that tally inconsistencies undermine citizen trust in the electoral authorities as impartial arbiters of elections.

What explains variation in the quality of hand-counted vote tallies? Making use of various electoral rules and procedures to identify causality, we find that more-educated poll workers yield tallies with fewer inconsistencies. An additional year of average educational attainment for the poll worker team associated with a polling station (4 to 6 citizens) reduces the extent of inconsistencies in the tally by up to 7%. The arithmetic difficulty of the tallying task, in contrast, renders inconsistencies more likely: the incidence of inconsistencies is about 17% greater when a sum in the *acta* requires carrying one than when it does not. Finally, mistakes are proportional to the workload, understood as the number of ballots cast, and therefore counted, in a given polling station. The incidence of inconsistencies increases by about 0.2% for every additional ballot cast.

[3] In concurrent elections there is more than one *acta* per polling station.

[4] On Mexico's long history of electoral manipulation before those reforms see Molinar (1991); Domínguez and McCann (1998); Cantú (2018).

A key contribution of this paper is to direct attention to the issue of the quality of vote tallies in 'normal' elections–i.e., clean, routine elections where votes are counted by people. The existing literature discusses the quality of tallies in two specific contexts: fraudulent elections (Myagkov et al., 2010; Hyde, 2007; Mebane, 2010; Cantú, 2018) and studies of voting technology (e.g., electronic voting machines) (Alvarez et al., 2009; Ansolabehere and Reeves, 2004; Alvarez and Hall, 2008). Virtually no attention has been given to the issue of tally quality in the modal case: elections where fraud is not an important issue and votes are counted by hand. We show that the quality of vote tallies varies considerably even within a single country where the administration of elections is centralized, as is the case in Mexico.

A second set of contributions of our analysis is to provide causal evidence showing that the quality of vote tallies is systematically explained by socioeconomic and behavioral factors, and that low-quality tallies have major political consequences such as fostering recounts and undermining the public's trust in the electoral authorities. The importance of the accuracy of vote tallies is recognized by the electoral laws of many countries. We believe it deserves greater attention from researchers.

Third, our results connect with ongoing debates about voting technology in the growing literature on election science. For one thing, our findings underscore the importance of considering the effects of tallying technology and procedures on the accuracy of totals. In addition, our findings on the (in)accuracy of hand counting underscore the existence of trade-offs between the possibility of electronic hacking by outside actors, on the one hand, and the accuracy of "hacking-proof" hand counts, on the other (Dee, 2007; Posner, 2000).

Our analysis also connects with discussions of electoral justice by raising the specter of a double development and democracy curse: countries and regions with low levels of overall development are also more likely to experience low-quality vote tallies, low trust in election results and in democratic institutions, and partisan strife. Finally, considering the literature on elections more broadly –including both empirical and game-theoretic work– our findings call into question the common assumption that converting cast votes into vote totals is a friction-less process, even in the absence of electoral malfeasance.

2 Related literature

A large fraction of existing research on the quality of vote tallies studies the relative performance of different voting and tallying technologies (Allers and Kooreman, 2009; Alvarez and Hall, 2008; Alvarez et al., 2009; Ansolabehere and Reeves, 2004; Lott, 2009; Dee, 2007; Garner and Spolaore, 2005; Mebane, 2004). We take technology (hand counting) as given, and instead focus on sociodemographic, workload, and cognitive causes of variation in tally quality. A few papers investigate sociodemographic correlates of voting technologies (Garner and Spolaore, 2005; Lott, 2009) and poll worker performance (Atkeson et al., 2014), but not of tally quality. We know of no research on the downstream electoral/political consequences of variation in tally quality, an issue we investigate.

The election forensics approach, meanwhile, tests for statistical fingerprints of fraud in official voting results (Beber and Scacco, 2012; Cantú, 2014; Kobak et al., 2016; Mebane, 2010; Myagkov et al., 2010). The fingerprints of fraud often correlate strongly with patterns of partisan advantage/disadvantage. Myagkov et al. (2010), for example, find that precinct-level turnout correlates with percent vote for the incumbent party, but does not correlate with votes for opposition parties. The inconsistencies we study, in contrast, bear no association to partisan performance and are not properly viewed as fingerprints of partisan malfeasance.

Research on trust in electoral results/institutions studies its relationship with voting technology (Alvarez et al., 2009), electoral fraud (Wellman et al., 2018), or individual traits (Alvarez et al., 2004), but not with the quality of tallies. Finally, our study joins a few recent data-intensive studies of Mexican elections, including Cantú (2018) on the fraudulent 1988 Mexican elections, Larreguy et al. (2016) on partisan monitoring of vote-buying brokers, Ascencio and Rueda (2017) on the effect of party representatives at the polls on electoral results, and Cantú and Ley (2017) on determinants of citizen participation as poll workers. None of these examine the quality of vote tallies.

3 Context: The counting of votes in Mexican elections

Mexico experienced electoral authoritarian government for most of the 20th century. After a series of crises, in the 1990s the major political parties negotiated a set of profound reforms to the electoral system that turned Mexico's regime into a democracy. The reformed system was designed to render partisan manipulation of elections very difficult. Its features included a transparent and reliable list of registered voters, a highly regulated process to select citizens to function as poll workers responsible for counting votes, a procedure to aggregate voting results quickly after polls close, a system of public financing and campaign spending rules that govern electoral campaigns, and an independent electoral tribunal of last resort to resolve electoral controversies. Perhaps chief among the reforms was the creation of an independent bureaucracy charged with organizing elections and producing official electoral results—the *Instituto Federal Electoral* (now called the *Instituto Nacional Electoral* or INE). Previously, all aspects of elections had been under the direct control of the executive branch of government.

3.1 Precincts (*Secciones*) and polling stations

The basic unit of Mexico's electoral geography is the *sección electoral* (subsequently precinct). Every precinct contains one or more polling stations (henceforth PS), depending on the number of voters registered in the precinct. To provide a sense for the magnitudes: there were 62,692 precincts and 129,238 PS in the 2012 presidential election. The average precinct covers about 1,200 registered voters. A strictly-enforced maximum of 750 registered voters can be assigned to vote at any given PS. This maximum determines the total number of polling stations needed in an election. Registered citizens are apportioned equally across the PS in a precinct. For example, in a precinct containing 752 citizens, two PS will exist, with 376 citizens assigned to each. The first PS in a precinct is known as the *básica* (basic) PS, the second PS is called *contigua 1* (contiguous 1), the third is *contigua 2*, and so on.

3.2 Assignment of citizen poll workers to polling stations

Mexican law requires that votes be tallied by randomly-selected citizens who function as polling-station workers (henceforth PW). Each PS is allocated 4 acting PW and 3 substitute PW.[5] INE recruits PW from the same area where the corresponding PS is located following a very detailed, transparent, and rigorous invitation procedure enshrined in the law. This procedure involves inviting a random set of 13% of registered voters in every precinct to function as PW. To achieve this, the INE hires a large team of professional recruiters to visit citizens at their home and assess their eligibility to function as PW. To be eligible, a citizen must not work for a political party and must be able to read and write, among other things. A second lottery then selects a randomly-chosen subset of the eligible citizens in a precinct, and these are designated to staff each of the PS in that precinct.

The assignment of designated citizens to PS within a precinct proceeds according to educational attainment. The citizen with the highest educational attainment is designated President of the first PS. The one with the next highest educational attainment is designated President of the second PS. Once every PS in the precinct has a President, the person with the next highest educational attainment is designated Secretary of the first PS. In the same manner, citizens are next designated to the positions of First Counter (*primer escrutador*) and Second Counter (*segundo escrutador*). The remaining citizens are designated as first, second, and third substitutes (*suplentes*). This assignment rule implies that the average educational attainment of poll workers is generally higher in the *básica* PS as in the *contigua 1* PS in the same precinct, a fact that we exploit further below to identify the causal effect

[5] Substitutes are also trained in case one of the four acting PW drops out or fails to show up on Election Day. When local elections are concurrent with national ones, the number of PW allocated to a PS increases to 6 acting and 3 substitutes.

of education on inconsistencies.

The general functions of the PW team for a given PS are to staff the PS during Election Day, to make sure only those eligible to vote at the PS do so, to count the votes by hand after the close of voting, and to fill out the *acta* that same evening. PWs are trained by thousands of INE employees (*capacitadores-asistentes electorales* or CAE) hired for that purpose.

3.3 Political party representatives

Political parties are entitled to send representatives to sit at the PS along with PW. These representatives are registered with INE by the political parties prior to the election. In general they are registered for a specific PS. They can observe the work of the PW, but they have no formal role in the ballot counting or in the filling out of the *actas*. There were 600,743, 743,263, and 846,336 party representatives in the elections of 2009, 2012 and 2015, respectively.

4 Data

We use seven sources of data. The main dataset contains measures of inconsistencies for each PS in the elections of 2009, 2012 and 2015. A second dataset describes the individual citizens who staffed each PS. A third one contains information on official aggregate vote results for every political party at the polling station level for each election. A fourth data source describes recounts at the polling station level. A fifth one documents the presence of political-party representatives at the polling station level. A sixth one is our survey to 80,000 PWs on attitudes towards INE. Finally, we use comprehensive socio-demographic data from the 2010 Population Census at the precinct level.[6]

Table 1 displays summary statistics. Panel A describes the number of precincts, polling stations, poll workers, registered voters, and votes cast for each of the elections of 2009, 2012 and 2015. It provides a glimpse of the breadth of the data we use. For each election our data cover approximately to 60,000 precincts, about 125,000 PS, half a million PW, more than 70 million registered voters, and between 31 million and 45 million votes cast, depending on the election. About 40,000 precincts contained more than one PS.

[6]We had no access to personally identifiable information such as voter names for any dataset.

Table 1: Summary Statistics by Election

	2009		2012						2015	
	Congressional		Congressional		Presidential		Senatorial		Congressional	
Panel A: Election-level variables										
Num. precincts	61,089				62,692				58,797	
Num. PS	126,198				129,238				123,319	
Num. registered voters	71,319,536				72,925,360				70,111,928	
Num. precincts with contiguous PS	42,136				41,523				39,184	
Num. PW	499,489				514,742				623,714	
Num. votes cast	31,671,852		45,584,376		45,577,568		45,572,336		33,368,108	
Turnout (%)	44.41		62.51		62.50		62.5		47.59	
Panel B: Inconsistencies in actas										
Num. inconsistency 1	13.4	(64.9)	7.2	(44.9)	8.1	(47.8)	7.3	(45.3)	6.9	(45.9)
Num. inconsistency 2	17.6	(74)	8.5	(44.6)	10.2	(49.2)	8.6	(44.6)	10.4	(54.4)
Num. inconsistency 3	2.4	(24.6)	2.8	(25.7)	2.9	(26.1)	3.0	(27.1)	2.7	(26.5)
Num. inconsistency 4	5.4	(34.8)	7.3	(37.5)	7.2	(38.5)	7.1	(37)	NA	NA
% of PS with incons. 1	12.5		7.7		9.0		7.7		8.0	
% of PS with incons. 2	24.9		32.0		32.7		32.0		27.6	
% of PS with incons. 3	4.4		7.5		8.1		7.7		6.9	
% of PS with incons. 4	29.1		38.2		38.3		38.0		ND	
% of PS with at least one inconsist.	41.6		43.6		44.5		43.2		30.5	
Panel C: Poll worker traits										
Age	34.8	(7)			36.0	(7.1)			37.2	(7.1)
% male	42.7	(25.9)			42.4	(25.3)			41.0	(22.9)
Years of education	11.6	(2.7)			11.9	(2.5)			11.1	(2.7)

Figures in parentheses are standard deviations. See text for further details. Panel C describes traits of PW. For these, we compute the average value for all PW within a PS and then average across PS.

4.1 Data on inconsistencies

For internal purposes, INE collects data on various types of inconsistencies in the *actas*. This is a massive undertaking: for each of the tens of thousands of PS in every election, INE records which inconsistencies were found in the data. We observe four numerical measures of inconsistencies in vote tallying at the PS level for the universe of *actas* from the Mexican federal elections of 2009 (legislative, lower house), 2012 (executive and legislative, both houses), and 2015 (legislative, lower house). The following pieces of information constitute the building blocks for our measures of inconsistencies:

- PV (*personas que votaron*): Total number of votes cast as checked off by the PW on the official voter list for the PS.
- RPPV (*representantes de partidos políticos que votaron*): Total number of votes cast in the PS by official representatives of political parties. Party representatives can cast a vote even if they are not on the voter list for the PS.
- SV (*suma de votantes*): Total number of votes cast in the PS, computed by the PW as the sum of PV + RPPN.
- BSU (*boletas sacadas de las urnas*): Total number of ballots extracted from the ballot box.
- RV (*resultados de la votación*): The sum of subtotals of votes cast for each of the political parties on the ballot plus write-ins and null ballots.
- BS (*boletas sobrantes*): The number of ballots that remain unused at the end of Election Day.
- TBE (*total de boletas entregadas*): The total number of blank ballots provided to the PS before the voting began, computed as the number of voters in the official voter list for the PS plus two ballots for each of the political parties listed on the ballot (since up to two representatives for every party can cast their votes in a PS where they are not registered but work as observers).

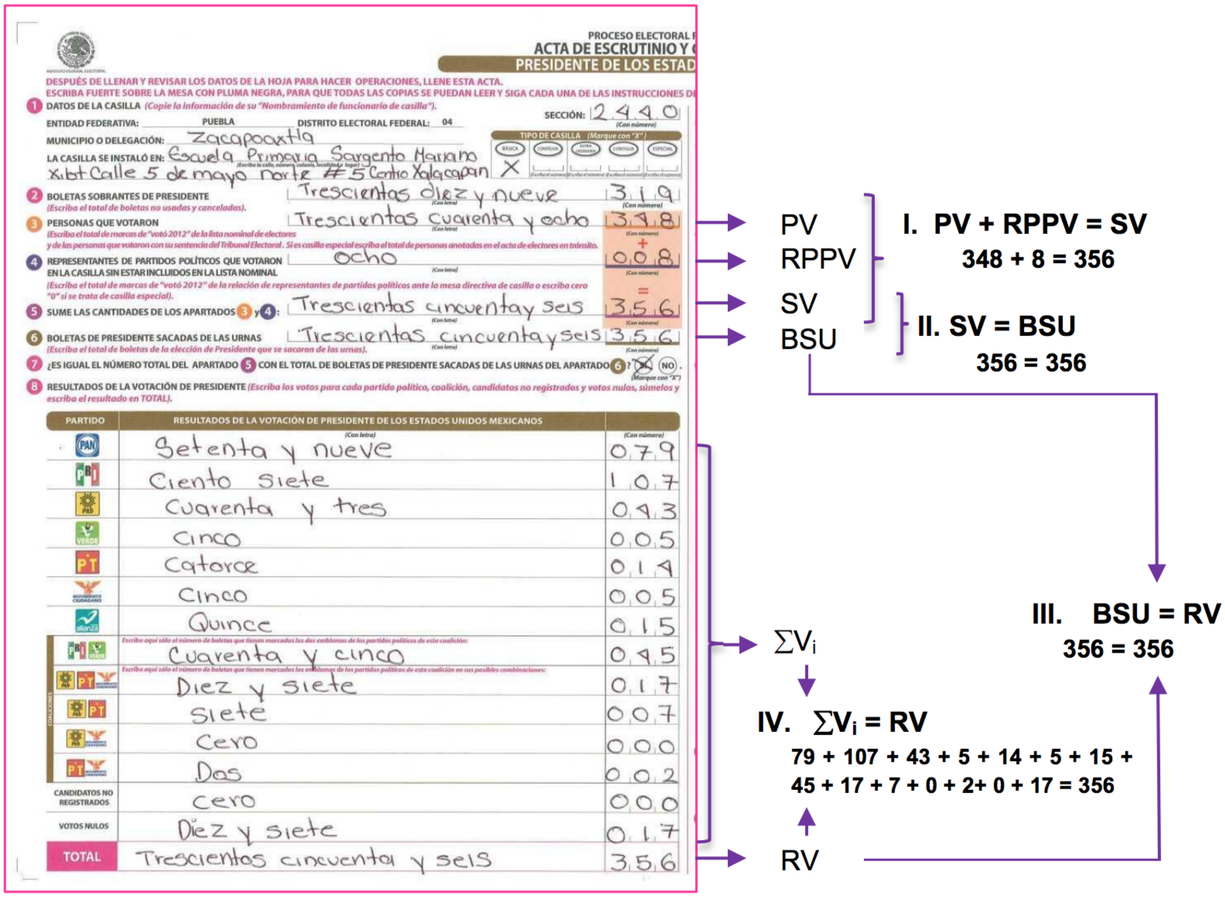
If an *acta* is filled out with no inconsistencies, the following equalities ought to hold:

1. SV = PV + RPPV (the sum of people who voted and party representatives who voted should be algebraically correct).
2. SV = BSU (the number of people and party representatives who voted should equal the number of ballots extracted from the ballot box).
3. RV = BSU (the sum of votes for parties, write-ins, and null ballots should equal the number of ballots extracted from the ballot box).
4. BS = TBE - BSU (the number of unused ballots should equal the total number of ballots provided minus the number of ballots extracted from the ballot box).

We define an inconsistency as a failure of one of the above equalities.[7] But given that they differ in nature –some are algebraic mistakes in the *acta* itself while others involve the actual number of ballots– we present results for each of them separately. We use the (absolute) magnitude of the discrepancy between the quantities at either side of an equality as our main measure of the degree of inconsistency. A sample *acta* fragment is shown below, illustrating the first three of our four measures.[8]

[7] These equalities were devised by INE and INE has utilized them at least since 2012 in order to internally describe

Figure 1: Sample *Acta* and Corresponding Inconsistency Measures



This figure shows part of an *acta* from the 2012 presidential election. Design varies slightly across elections. Item 3 corresponds to what we termed PV above, item 4 corresponds to RPPV, item 5 is the sum of 3 and 4 (SV), and item 6 corresponds to BSU. Item 8 displays the vote subtotals for each political party; the total of these corresponds to RV. The rightmost half of the figure illustrates the inconsistency measures we use. The *acta* has a signature page that is not displayed here. Physical images of the *actas* are available at http://siceef.ine.mx/.

4.2 Data on PW characteristics

For each PW we observe age, gender, and years of education completed. We use data on those PW who attended their polling station on Election Day.[9] Panel C of Table 1 displays polling-station averages. On average, PW are in their late thirties, close to 42 percent are male, and completed about 11 to 12 years of education—that is, almost completed high school. Variation across PS is significant, with standard deviations of 7 years of age, 2.7 years of education, and 25pp in percentage male.[10]

the quality of the *actas* in national elections. They have legal bite as these same inconsistencies can legally trigger recounts.

[8] We cannot use the equality labeled as IV in the figure because we do not have the line-by-line data on number of votes for each of the parties; only the total is available in machine readable form.

[9] Some fraction of PW do not show up at their PS on Election Day.

[10] We also have data on PW training. There is about one CAE for every 6 polling stations, on average. We observe which citizen PW was trained by which CAE. We also observe CAE characteristics including education level, score

4.3 Data on political party representatives

For elections in the years 2012 and 2015, we observe which party sent representatives to which PS. The main political parties—PAN, PRI and PRD—respectively sent monitors to 73%, 93%, and 55% of PS in 2012. The equivalent figures for 2015 are 81%, 94%, and 63%. Party representatives have to be registered at INE at prior to the election, and their identities are verified by PW on Election Day. Party representatives can vote in a PS even if they are not on the voter list for that PS or precinct.

4.4 Data on recounts

Our data indicate whether the votes in a given PS in a particular election were recounted, for all elections in 2009, 2012 and 2015. For a PS to be recounted a political party has to officially request a recount. The law provides as valid reasons to request a recount: i. The presence of inconsistencies on the *acta* that cannot easily and readily be corrected or explained away; ii. A margin of victory (at the *acta* level) smaller than the total number of null votes cast in the *acta*; and iii. The situation where all votes in the *acta* are for the same party. There were 34,795, 198,007, and 77,113 *actas* recounted in 2009, 2012 and 2015 respectively. These amount to 27.6%, 51.1%, and 62.5% of the total number of *actas* in the corresponding election years.

4.5 Data on voting results

Aggregate vote results are public information published by INE in the SICEEF (Federal Elections Statistical System of Queries). For each PS we observe the number of votes received by each party for the lower house of congress in 2009, 2012, and 2015, and also for president and the upper house of congress in 2012. From other data sources we have information on the number of registered voters as well as on the PS and precinct to which each registered voter was assigned to vote. We were able to verify that the rule that no PS can have more than 750 registered voters allocated to it is strictly followed.

4.6 Survey of poll workers

The survey was implemented in the 2017 local elections in Coahuila, Estado de Mexico, Nayarit and Veracruz. The survey was targeted to a random sample of almost 80,000 PW of the election and implemented by the CAE during the second stage of the PW recruitment

in the interview and job exams, and experience as CAE in previous elections. We use these data as controls in regressions.

process. Randomization was conducted at the level of ARE (an ARE is a set of contiguous precincts assigned to a CAE). Of the 6,690 CAE, 4,014 were selected to implement the survey to all the PW in the precincts under their charge. We collected 85,006 completed surveys in 7,161 different precincts. Figure A10 in the online Appendix shows the precincts in which at least one PW completed a survey.

4.7 Socioeconomic data

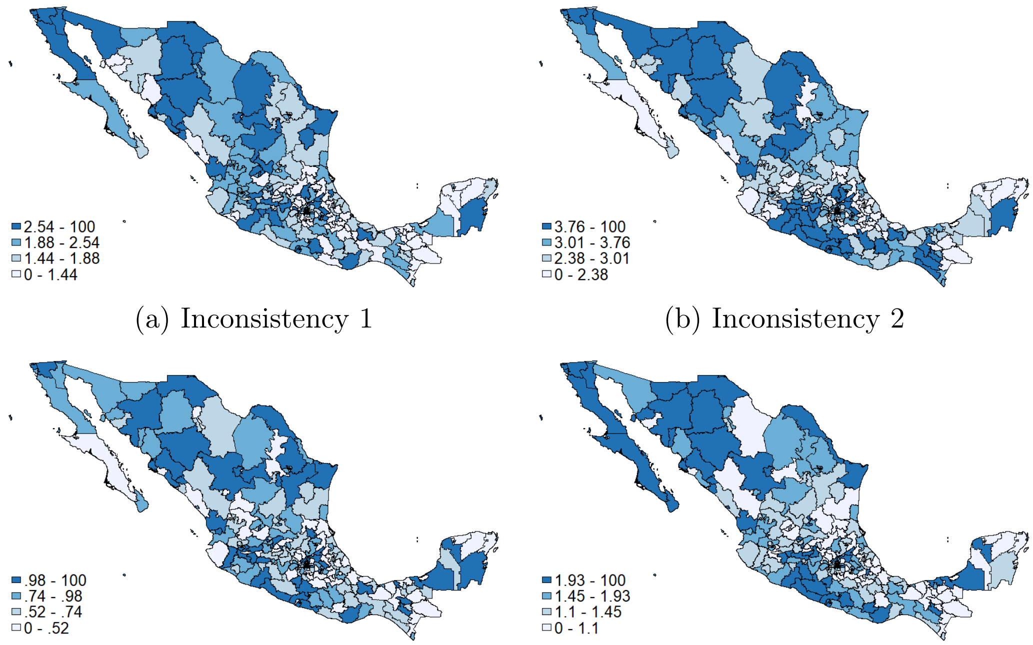
INE and INEGI provide a version of the most recent Population Census (2010) where data are presented at the precinct level. The data cover 66,740 precincts. From this dataset we use the following variables as controls in regressions: (i) percentage of the population constituted by: men, residents of a state born in the same state, indigenous, catholic, without social security, employed; (ii) average years of education, and (iii) percentage of households with: water, sewage, dirt floor, electricity, radio, TV, refrigerator, car, computer, telephone, cellphone, and internet. Averages across precincts (considering the full census) are: 1,075 voting-age adults, 8.23 years of education, 37.4% of the population employed, 42.6% of households with car, and 95.3% of households with electricity (all in per-precinct units).

5 Inconsistencies in vote tallies

Inconsistencies in vote tallies are prevalent, they are a nation-wide problem in Mexico, and they do not seem to be going away. Panel B in Table 1 provides descriptive statistics for the measures of inconsistency defined in Section 4.1. The first four lines provide the average (absolute) discrepancy between the two sides of the corresponding equality. This is a measure of the *extent* of inconsistency. For example, in the 2012 Presidential elections, equality 2 failed to hold by an average of 10.2 votes. We henceforth refer to this measure as "inconsistencies" unless indicated otherwise. The last four lines in Panel B describe the percentage of PS where the corresponding equality did not hold. This is a measure of the *presence* of inconsistencies that does not consider their extent. For example, equality 1 failed to hold in 9% of PS in the 2012 Presidential election. Type 2 and 4 inconsistencies are the most common, with around 25%-38% of PS displaying these.

As Panel B shows, there is no indication of temporal trends either in the overall magnitude or prevalence of inconsistencies at the country level. Figure 2 shows inconsistencies for each electoral district. Each color represents a quartile and darker colors indicate more inconsistencies. The main thing to note is that there are inconsistencies in almost every area of Mexico. There seems to be some geographical clustering, suggesting possibly common determinants.

Figure 2: Geographical Distribution of Inconsistencies



(a) Inconsistency 1

(b) Inconsistency 2

(c) Inconsistency 3

(d) Inconsistency 4

The maps display the geographical distribution of inconsistencies by electoral district, pooling together all election years and election types in our data. Specifically, for each inconsistency of type $j = 1, .., 4$, the figure plots the ratio of the absolute number of type-j inconsistencies ($AbsNumInc^j$) in the district over the total number of votes cast, multiplied by 100. Shading represents quartiles.

The different measures of inconsistencies are not strongly correlated (Table A2 in the Appendix). Types 1 and 2 are weakly correlated with types 3 and 4. The strong correlation (.75) between types 1 and 2 is likely due to the fact that they share the term SV (total number of votes cast), which is the result of a sum computed by the PW and therefore prone to mistakes.

6 Are inconsistencies partisan?

As mentioned previously, Mexico's electoral history makes it necessary to explore whether the inconsistencies we study reflect partisan manipulation or unintentional mistakes.[11] Some kinds of partisan fraud could in theory result in the kinds of inconsistencies we study. For example, if the cheating party stuffed ballot boxes with extra, pre-marked ballots, then the number of people checked off on the voter list (SV) would be smaller than the number of

[11] Crespo (2006), for example, argues that because the extent of inconsistencies in the *actas* in the 2006 presidential election exceeded the overall margin of victory, it is not possible to know who was the rightful winner. Others, however, question the validity of this claim on the basis of the evidence (Pliego Carrasco, 2007; Aparicio, 2009).

ballots in the ballot box (BSU), violating the second equality described in section 4.1. Many other kinds of electoral manipulation, however, would not result in inconsistencies in the tallies. These include padding the voter lists and tampering with the aggregate vote count. In today's Mexico, these forms of electoral manipulation have become the exception rather than the rule (Cantú, 2014). Electoral manipulation in today's Mexico takes primarily the form of vote buying and violations of campaign finance (Serra, 2016). While reprehensible and illegal, vote buying and campaign finance violations are not causes of inconsistencies in the tallies.

The fact that inconsistencies in the tallies are an important cause of recounts could give rise to a mixed set of incentives for political parties and their representatives at the PS. On the one hand, if a party were cheating in a particular PS it might wish to avoid inconsistencies in order to avoid scrutiny. On the other hand, a party that stood to lose in a given PS could benefit from inducing a recount (or, in the limit, an annulment), and therefore would benefit from creating inconsistencies in the tally.[12]

Political parties, however, have very limited means to influence whether or not a tally displays inconsistencies, because the tallying is done by nonmilitant citizen PW selected at random. Political parties have the right to send representatives to every PS to observe the tallying. These representatives have no formal authority with respect to the PW and the vote tally, but they could attempt to informally influence the PW team, for example, in decisions about whether a particular ballot was marked in a valid way or ought to be annulled. They could also check the tally and ask for the PW to resolve any inconsistencies—but there is no obvious way in which a representative could induce inconsistencies in the tally.[13]

To empirically investigate the possibility that inconsistencies might have partisan causes, we run the following set of analyses. First, we check the association between the fraction of the vote that goes to each of the political parties and the extent of inconsistencies. Second, we study the association between the presence of party representatives in a given PS and the extent of inconsistencies. Third, we check whether the extent of inconsistencies in a given PS persists over time through different elections. This last analysis explores the possibility that the influence of political parties on inconsistencies depends on the local organizational capabilities (the "machine") of the parties, which should ostensibly persist over the period of

[12] The annulment of a full polling station is rare, but one cause of annulment is the presence of mistakes in the vote tally (*Ley General del Sistema de Medios de Impugnación en Materia Electoral*, article 76, `http://www.diputados.gob.mx/LeyesBiblio/pdf/149_190118.pdf`). In the 2012 election, for example, only 526 out of 143,132 PS (about .36%) were annulled (`http://portales.te.gob.mx/`).

[13] Ascencio and Rueda (2017) find that the presence of party representatives is associated with small increases in the corresponding party's vote share in legislative elections in Mexico. We regard their finding as compatible with ours: there is no obvious reason why efforts by party representatives to influence whether a ballot is regarded as null or as favoring a party, for example, need also influence the consistency between the different fields of a vote tally (such as the sum underlying our first indicator of tally quality). Indeed, as we report below, we find no association between the presence of party representatives and tally quality.

time in our data.

The full details of these analyses are described in the online Appendix for reasons of space. We find that the estimated association between the party vote and inconsistencies is substantively tiny. For example, an additional 500 inconsistencies of type 1, or 2,401 fewer inconsistencies of type 3, would be required to "generate" one additional vote for the PRD. These estimates make it difficult to argue that inconsistencies are associated with substantively-important misallocation of votes to parties. We next find that the presence of political-party representatives at the PS is not an important correlate of inconsistencies. The presence of a PAN representative, for example, is associated with an additional .21 inconsistencies of type 1 or .25 of type 3, but neither estimate is statistically significant. We also find that the presence of party representatives does not moderate the relationship between inconsistencies and the partisan vote. Finally, we find that inconsistencies do not persist over time within precincts. Overall, these results, together with the scholarly consensus on the state of contemporary Mexican elections, suggest that inconsistencies do not arise out of partisan manipulation, but instead primarily reflect honest mistakes.

7 Causes of inconsistencies in vote tallies

The evidence so far suggests that inconsistencies in vote tallies do not reflect malfeasance. We therefore look for causes of inconsistencies in factors that could plausibly drive honest mistakes when tallying votes. We find clear causal evidence that the educational attainment of those selected as poll workers, the difficulty of the tallying task, and the workload of the PW increase the incidence of inconsistencies in vote tallies.

7.1 Education

Does the quality of vote tallies depend on the education/numeracy of the citizens selected to function as PW? Ex-ante, we are agnostic on this point. Lower educational attainment could make it more difficult for poll workers to successfully complete their tallying tasks without mistakes. At the same time, anecdotal evidence suggests that PW with lower educational attainment take their vote-tallying tasks more seriously and therefore exert greater effort than their more-educated peers.

Simply regressing the extent of inconsistencies on the educational attainment of PW could potentially be subject to concerns about omitted variable bias. To mitigate this possibility, because the average level of education in the population is likely differ across precincts, we focus on variation in educational attainment across PS *within* a precinct. In addition, we use an exogenous source of within-precinct variation in the educational attainment of PW based on the procedure used to allocate PW across PS.

As described earlier in the paper, using a random selection procedure INE selects a pool of eligible and willing PW for every precinct. Whenever there is more than one PS in a precinct, these PW are allocated to the various PS within their precinct according to the following rule: The person with the highest educational attainment is named President of the first PS; the second most-educated person is named President of the second PS; etc. Once all PS have a President, the next most-highly educated person in the pool is assigned to be Secretary of the first PS; the next one is named Secretary of the second PS; and so forth, until every PS has a full set of PW (either four or six PW, depending on the number of concurrent elections).[14]

This allocation rule has the consequence that the team of PW assigned to the first polling station (*básica*) in a precinct has a higher level of educational attainment on average than the PW team assigned to the second polling station (*contigua 1*), which in turn has higher average education than the team assigned to the third polling station (*contigua 2*), etc. Figure A2 in the online Appendix shows that indeed PS are ranked by education. Panel (b) in that figure shows that on average, PW working at first-ranked polling stations have about 0.6 more years of education than second-ranked polling stations, 0.81 more than third-ranked polling stations, and 0.93 more than fourth-ranked polling stations. Crucially, this variation is entirely due to the allocation rule, and is therefore plausibly orthogonal to potentially-confounding traits of the polling stations. On this basis we implement the following instrumental-variables strategy:

$$AbsNumInc^{j}_{pste} = X'_{pst}\alpha + \beta^{j} S_{pst} + n_{st} + u_{pste} \tag{1}$$

$$S_{pst} = X'_{pst}\pi_0 + \mathbb{1}(B_{pst})\pi_1 + n_s + e_{pst} \tag{2}$$

where S_{pst} is the average years of educational attainment of the PW for PS p in precinct s in election-year t; X_{pst} is a matrix of covariates that includes the average age and the fraction who are female of the PW team for PS p in precinct s in election year t, as well as various traits of the recruiter (CAE) who recruited and trained the PW in all PS in precinct s;[15] and $\mathbb{1}(B_{pst})$ is an indicator that PS p in precinct s in election-year t is the first PS (*básica*). The coefficients of interest are the β^{j}.

Table A8 in the online Appendix presents results for the first-stage regression. The first

[14] For a sample allocation and the rule itself see `https://tinyurl.com/ya6h3g67`.

[15] These include the age, gender, educational attainment, and hiring-test score of the CAE.

stage is very strong, as expected, with $\pi_1 = 0.897$, a t-stat above 500, and an F-stat of 50,968. Table 2 displays the second-stage estimates. An additional year of average education in a PS reduces the absolute number of inconsistencies of type 1 by 0.5, of type 2 by 0.66, of type 3 by .03, and of type 4 by 0.3. These represent 6%, 7%, 1% and 4% of the corresponding means. All of these estimates are statistically significant (except in the case of type-3 inconsistencies). These results imply that selecting PW with greater educational attainment would result in vote tallies with fewer inconsistencies.[16]

Table 2: Effect of Poll-Workers' Education on Tally Quality: IV Estimates

	(1) Inconsistency 1	(2) Inconsistency 2	(3) Inconsistency 3	(4) Inconsistency 4
Years of education	-0.503**	-0.657***	-0.033	-0.299**
	(0.21)	(0.21)	(0.10)	(0.15)
Age	0.031	0.052**	-0.007	-0.000
	(0.03)	(0.02)	(0.01)	(0.02)
Male (%)	0.315	-0.662	0.436	-0.045
	(0.61)	(0.59)	(0.29)	(0.44)
Constant	-0.345	44.001**	1.689	-8.137
	(16.76)	(17.38)	(8.38)	(14.51)
N	368242	436398	433594	354923
R-sq	0.00	0.00	0.00	0.00
Mean of the inconsistency	8.6	11.3	2.8	6.8
Education mean	10.1	9.9	9.9	9.9
Controls	Yes	Yes	Yes	Yes
Precinct FE	Yes	Yes	Yes	Yes

The analysis excludes precincts containing only one PS. We also exclude 994 PS located in atypical precincts containing more than 14 PS each. *** P<.01; ** P<.05; * P<.1.

7.2 Difficulty

Insofar as education, numeracy, or training matter for the quality of vote tallying, one would expect that more-difficult tallies should on average exhibit more inconsistencies than easier ones. To explore this possibility, we construct a measure of the difficulty of the tallying task. One natural measure of tallying difficulty is the arithmetic difficulty of a sum. The first type of inconsistency requires that PW perform a sum. The sum generally involves a

[16] Although we believe it plausible that education itself might help to develop skills helpful to tally votes without making mistakes (e.g., in arithmetic), our analysis estimates the effect of selecting people with greater educational attainment as PW —not the effect of marginally increasing the educational attainment of PW keeping all else constant. Because educational attainment is associated with a variety of other factors in the population, we include controls for the gender an age of PW, as well as precinct fixed effects.

"large" number (i.e., the number of votes cast in a PS, which is usually in the hundreds) and a "small" one (i.e., the number of party representatives who cast votes in the PS, usually smaller than 10). We classify such a sum as "difficult" if it involves carrying one over, and as "easy" if it does not.[17] We construct a dummy variable that takes the value of 1 when the sum that a PW needs to perform is "difficult" and the value of 0 when it is "easy." Close to 35% of the tallies contain difficult sums.

We believe the difficulty of the sum, thus defined, can be regarded as exogenous with respect to inconsistencies in the tally. For one thing, it depends to a large extent on the last digit of the total number of votes cast in a PS. Crucially, whether turnout is low or high should have no bearing on the last digit of the total number of voters. Still, we check for balance on observables between those tallies where the sum in question is difficult vs. those where it is easy. Table 3 presents the results of the balance tests. Each of the first four columns represents the regression of a predetermined covariate on the difficulty dummy. These covariates are: a dummy indicating whether the PS is the first one (*básica*) in the precinct or not, the average years of education of PW in the PS, the share of male PW within the team at the PS, and the average age of the PW team at the PS. As before, the estimates are based on variation across PS within a precinct (i.e., they include precinct fixed effects). As expected, there is no difference in any of the covariates between PS with a difficult vs. an easy sum.

Column 5 displays the effect of the difficulty indicator on the extent of inconsistencies of type 1 (the type that involves the aforementioned sum). A difficult sum, in comparison with an easy one, increases the extent of inconsistencies by 1.46, that is, by 17% of the average extent of inconsistencies of type 1 (equal to 8.58). To further probe whether our measure of difficulty indeed relates to the kinds of skills that presumably correlate with formal education, we study whether the effect of difficulty on inconsistencies is moderated by the education of the PW. In column 6 we interact the difficulty dummy with the average educational attainment of the PW team in the relevant PS. As before, the main effect of average education is negative. The effect of difficulty, however, is a function of education. Every additional year of educational attainment reduces the effect of difficulty on the extent of inconsistencies by .31. The coefficient on the difficulty dummy is 5.28, implying that the effect of difficulty on inconsistencies is completely nullified when the average level of educational attainment among PS workers is about 17 years.

[17] For example, 234+2 does not require carry over, but 234+8 does.

Table 3: Effect of Tallying Difficulty on Tally Quality

	(1) Basica	(2) Education	(3) Male (%)	(4) Age	(5) Incon. 1	(6) Incon. 1
Difficulty dummy	-0.003	-0.006	0.000	0.008	1.457***	5.284***
	(0.00)	(0.01)	(0.00)	(0.05)	(0.33)	(1.77)
Education						-0.400***
						(0.12)
Difficulty * Education						-0.314**
						(0.14)
Constant	0.481***	11.917***	0.420***	36.216***	8.076***	12.838***
	(0.00)	(0.00)	(0.00)	(0.02)	(0.11)	(1.40)
N	478752	478649	478707	478628	472834	472736
R-sq	0.345	0.854	0.557	0.651	0.447	0.447
Y mean	0.480	11.915	0.420	36.219	8.577	8.575
Difficulty mean	0.344	0.344	0.344	0.344	0.344	0.344
Precinct FE	Yes	Yes	Yes	Yes	Yes	Yes

Columns 1-4 display balance tests, where a predetermined covariate is regressed on the difficulty dummy: $Predetermined_{pste} = \alpha + \gamma DifficultyDummy_{pste} + n_{st} + \epsilon_{pste}$. Such covariates include: indicator for whether the PS is first-ranked, average education of PW at the PS, fraction male PW at the PS, average PW age at the PS, and trainer (CAE) evaluation score. Variables are indexed by: PS p, precinct s, election year t, and election-type e. The unit of observation is an *acta*. In columns 5-6 the dependent variable is $AbsNumInc^1_{pste}$, the absolute number of inconsistencies of type 1. Standard errors clustered at the precinct-year level shown in parentheses below coefficient estimates. *** P<.01; ** P<.05; * P<.1.

7.3 Workload

The final cause of tally quality that we test is the workload faced by poll workers. The issue of workload figures prominently in current debates in Mexico. On election day, a PW spends about 12 hours staffing and managing her assigned precinct, and then about 3 additional hours tallying up the votes and filling out the *actas*. INE is concerned that excessive workload could lower the quality of the vote tallies.[18] They may be justified: a large literature in psychology and neuroscience shows that attention, self-control, and cognitive function in general are subject to fatigue through mechanisms such as glucose depletion.[19] In fact, the rule that polling stations should have no more than 750 voters was motivated by the desire to limit workload and reduce PW mistakes, and INE is considering implementing electronic voting to further reduce the burden on PW.[20] Academics and policy makers have similarly used a workload argument to support electronic voting,[21] but unfortunately there seems to

[18] https://tinyurl.com/yb33j4fa, and https://tinyurl.com/ycjkzd5a.
[19] https://tinyurl.com/y7ltx5r8
[20] The issue has gained even more relevance now since INE has acquired authority over the management of local elections, which implies that the same citizen PW now have to count the ballots for both federal and local elections when these take place concurrently.
[21] e.g. https://tinyurl.com/ybers5pj.

exist no quantitative evidence for or against the workload conjecture.

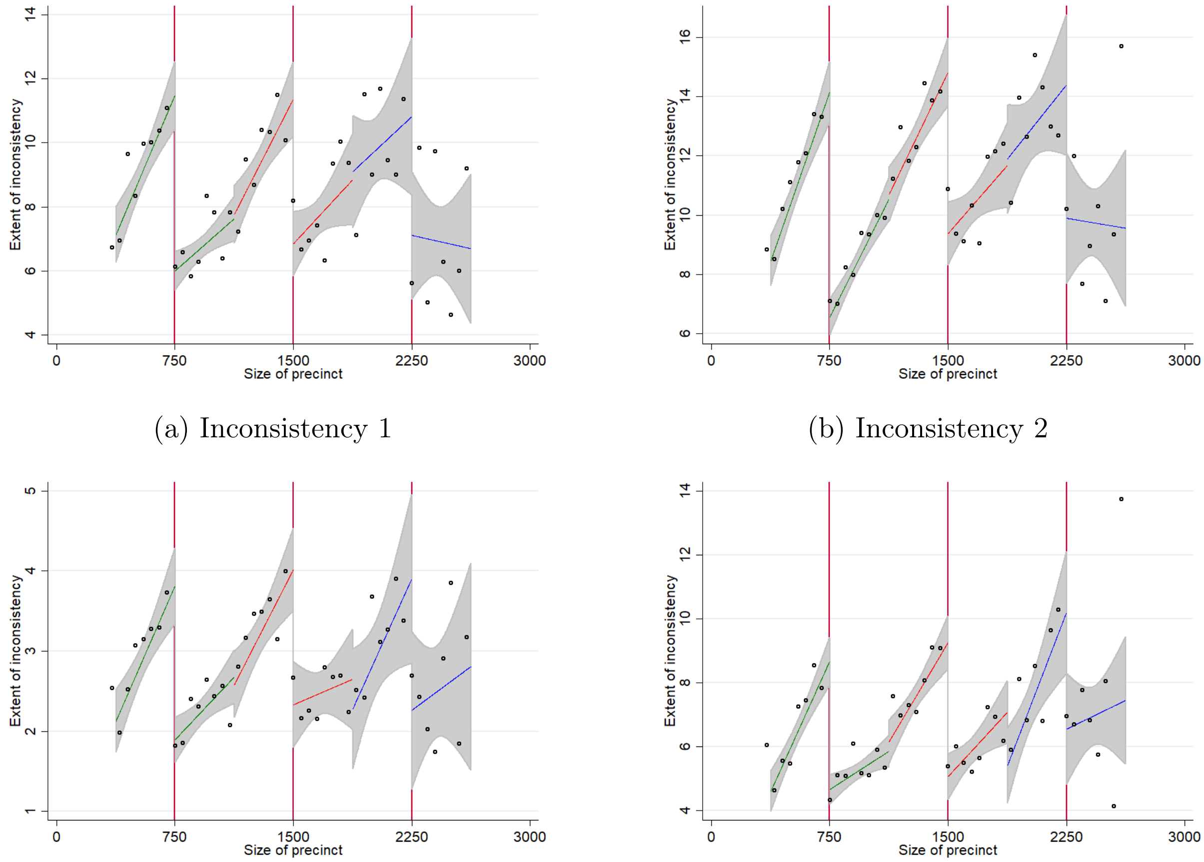
The ideal experiment to test the workload hypothesis would allocate more (or fewer) voters randomly to some polling stations, and measure how this translates into more or fewer mistakes in the tally for the PS. We approximate the notional experiment through a natural experiment. Specifically, we exploit the previously-mentioned fact that precincts are capped at 750 registered voters by law. If the number of registered voters in a precinct exceeds 750, an additional PS is added and the voters are apportioned equally across all the PS in that precinct. This rule, therefore, generates a discontinuity in the number of registered voters assigned to each PS at precinct sizes that are multiples of 750 where we can apply a regression discontinuity (RD) methodology. For instance, a precinct with 750 registered voters only has one PS, while a precinct with 751 registered voters has two PS, respectively with 375 and 376 registered voters each. This legal cap on PS size is followed quite strictly (online Appendix Figure A3).

To estimate the causal effect of workload on the extent of inconsistencies, we implement a regression discontinuity analysis. In the online Appendix we present McCrary density tests (Figure A4) and smoothness/balance tests (Figure A5 and Table A9), which lend strong support to the RD identification assumption. Figure 3 presents the main results graphically, separately for each of the four types of inconsistencies. The horizontal axes describe the number of registered voters in a precinct, while the vertical axes represent the extent (absolute number) of inconsistencies (one inconsistency-type is shown in each panel). The vertical lines indicate the number of registered voters at which an additional PS is to be added, inducing the jump in the number of registered voters per PS in the precinct that we use to identify causality. The regression estimates corresponding to the figure are provided in Table A10 in the online Appendix.

The pattern that emerges from the figure is quite clear: workload—the number of ballots to be tallied— causes inconsistencies. The figure shows, for example, that number of inconsistencies is halved at 751 registered voters, and again decreases sharply at the 1501 registered voters. In between the discontinuity points, the slope (inconsistencies per registered voter) is positive and practically linear. This pattern is present for each of the four types of inconsistencies. Respectively, the extent of inconsistencies for types 1,2,3, and 4 decrease by 5.5, 7.6, 1.9, and 4.0 at the 751 discontinuity (the mean extent of inconsistencies just below the 751 cutoff is roughly 10, 14, 5, and 7 for each of types 1-4, respectively). These are substantial decreases and they are all precisely estimated (with t-stats above 5).

Generally speaking, one might postulate two simple models of inconsistencies as a function of workload. The first is simply that each vote has some probability of being erroneously tallied, independently of how many votes have been counted before it. This model would

Figure 3: Effect of Workload on Tally Quality (Regression Discontinuity Analysis)



(a) Inconsistency 1

(b) Inconsistency 2

(c) Inconsistency 3

(d) Inconsistency 4

The dots denote bins (30 point width) and the trendlines are linear fits with shading representing 95% confidence intervals. See text for further details

imply that the level of mistakes increases proportionally to the workload (i.e., to the number of votes counted). A second model, consistent with fatigue explanations, is that mistakes are a convex (instead of linear) function of total votes. In this case, the likelihood that an additional vote is miscounted would increase with the number of votes counted previously by the PW team on election night. This distinction has important policy implications. In the second model, further reductions in the number of registered voters per PS—a measure that INE has considered—would reduce the extent of inconsistencies, but this would not be true in the first model.

Figure 3 suggests that the relationship between workload and inconsistencies is in fact linear. To investigate this, we redefine the dependent variable as the ratio of the extent of inconsistencies over the workload (number of votes counted) in a PS. We find (online Appendix Figure A6) that the slope is flat and there is no jump in inconsistencies at the discontinuity points—that is, the rate of inconsistencies per vote counted is approximately constant.[22]

8 Consequences of Inconsistencies

To be sure, mistakes in vote tallies, even if nonpartisan in nature, violate basic tenets of democratic fairness and are therefore undesirable. But inconsistencies in tallies also have serious practical consequences. For one thing, they can be, and often are, used by politicians to undercut the legitimacy of an electoral result or of democratic institutions themselves. We indeed find that inconsistencies in vote tallies make recounts substantially more likely, and that in doing so they erode public trust in the electoral authorities. Additionally, in the online Appendix, we present evidence that in the tightest races, inconsistencies could potentially deprive the rightful winner of their victory.

8.1 Tally quality and ballot recounts

As mentioned previously, inconsistent vote tallies are a legal reason for recounting ballots in Mexico and many other countries (we provide a partial list in Table A11 in the online Appendix). They are also political ammunition often used to question election results. To study the relationship between inconsistencies and recounts, we create a dummy variable indicating whether a PS was subject to a recount. The share of PS subject to a recount ranges in our data between 27% (in the 2009 legislative elections) and 62% (in the 2015 legislative elections). The mean over the five national elections in our data is 48.6%. We estimate the relationship between the presence of inconsistencies and the likelihood of a recount using the following linear probability model:

$$\mathbb{1}(PS\_Recounted)_{pste} = \alpha + \beta^j \mathbb{1}(AbsNumInc^j_{pste} > 0) + n_{st} + \epsilon_{pste} \tag{3}$$

where $\mathbb{1}(PS\_Recounted)_{pste}$ is a dummy variable indicating that PS p in precinct s in year t and election-type e was recounted and $\mathbb{1}(AbsNumInc^j_{pste} > 0)$ is a dummy variable indicating that the number of inconsistencies of type $j = 1, .., 4$ in absolute value was greater than zero. For this analysis, we use this indicator of the presence of inconsistencies, instead of a measure of their extent, because it is their presence that the law marks as a cause for

[22] We observe similar results when the dependent variable is vote-counting time (Figure A7 in the online Appendix). Under the fatigue hypothesis one might have expected, in contrast with this finding, that time should be a convex function of the number of votes counted. It is conceivable that fatigue should become an important driver of inconsistencies if workloads were greater than 750, but since no one PS has more than 750 voters we cannot test this.

requesting a recount.[23] The unit of observation is an *acta*. We include for precinct-year fixed effects (n_{st}) to control for location-specific variables like education, socioeconomic status of the neighborhood, and local strength of the political parties, among other factors.

Table 4 presents the results. Columns 1-4 show that the presence of each of the four types of inconsistency is individually strongly related to the likelihood of recount, with effect sizes ranging between 8.8pp and 26pp. Column 5 shows that the presence of any inconsistency in the vote tally is associated with a 22.5pp greater probability of a recount.

Table 4: Tally Quality and Recounts (OLS and IV Estimates)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) IV
Inconsistency 1	0.088***					
	(0.00)					
Inconsistency 2		0.262***				
		(0.00)				
Inconsistency 3			0.161***			
			(0.00)			
Inconsistency 4				0.170***		
				(0.00)		
Any inconsistencies					0.226***	0.679***
					(0.00)	(0.10)
Constant	0.447***	0.362***	0.426***	0.360***	0.347***	0.147***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)
N	472974	561608	556592	459021	587998	587998
R-sq	0.572	0.582	0.563	0.530	0.575	0.0844
F-stat						94.58
Precinct FE	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variables in columns 1-4 are each of the four types of tally inconsistencies, respectively. In columns 5-6, the dependent variable is a dummy variable that takes the value of 1 if any inconsistency of any of the four types was present. Column 6 instruments $\mathbb{1}(AbsNumInc_{pste}^{j} > 0)$ with an indicator variable for whether the PS is the first-ranked within its precinct. Standard errors shown in parentheses below coefficient estimates. *** P<.01; ** P<.05; * P<.1.

We emphasize that these estimates are identified on the basis of variation between different PS *within one same precinct*. In other words, if the vote tally for one PS displays no inconsistencies and the tally for another PS in the same precinct does, the latter is about 22pp more likely to be recounted than the former. In order to give a causal interpretation to the regression estimates, it is sufficient to assume that the various PS within a precinct would have had the same probability of being recounted if none had displayed inconsisten-

[23] Consistent with this, Figure A9 shows the presence of inconsistencies is an important determinant of recounts but their the extent less so.

cies. We believe this is a reasonable assumption in light of the fact that the PS in a precinct are generally located in the same physical space (e.g., a school building), and precincts cover a narrow geographical space.

Nevertheless, we additionally implement an instrumental variables strategy. We instrument for inconsistencies with the allocation rule that determines which poll workers are assigned to which PS within a precinct. In section 7.1 above we showed that a dummy variable indicating whether a PS is the first one in the precinct (*básica*) predicts inconsistencies.[24] The identifying assumption is that, within a precinct, this dummy variable may only cause recounts via its effect on inconsistencies. We find no reason to believe otherwise. Column 6 of Table 4 presents the instrumental variables estimator. The result that inconsistencies cause recounts stands. In fact, the IV point estimate is larger than the comparable one based on the OLS regression (column 5).[25] The result is also robust for controlling for other major causes of recounts as stated in the law.[26] In sum, the analysis furnishes evidence that inconsistencies in the vote tallies are an important cause of recounts.

8.2 Inconsistencies and trust in the electoral authority

"Inconsistencies make it easy to sow doubts about elections and difficult to clear such doubts," writes Schedler (2009). Once they get into the public eye, inconsistencies in vote tallies can undermine trust in election outcomes and in the electoral system itself—often with the help of political rhetoric. Media coverage of inconsistencies typically takes place in the context of partisan calls for recounts and of the recount processes themselves. In this section, we explore the relationship between inconsistencies in vote tallies and trust in the electoral authorities. We measure this trust variable through an original survey of over 80,000 Mexican citizens conducted in 2017 in the states of Estado de México, Veracruz, Coahuila, and Nayarit.[27] Respondents are asked to state the extent to which they agree with the statement that "INE is impartial and does not favor any political party." Answer options consist of a five-point scale: strongly agree (=5), agree (=4), neither agree nor disagree (=3), disagree (=2), and strongly disagree (=1).

Having shown in the previous section that inconsistencies are an important cause of recounts, we now focus on the relationship between recounts and trust in the electoral authorities. Because even with 80,000 surveys the trust data are too sparse to compute

[24] This relationship is now the first stage in a 2SLS instrumental variables estimation (the F-stat is 94.5).

[25] The two estimates are not directly comparable: the IV estimates a LATE, while the OLS estimates an ATE.

[26] Removing all PS that meet the aforementioned legal criteria for recounts other than the presence of inconsistencies does not affect the estimates.

[27] The survey was implemented during the second stage of the process through which INE recruits citizens to function as PW. The sampling frame was a random sample of all PW recruited to staff PS in the 2017 state elections. Section 9 in the online Appendix describes the survey and its coverage.

polling-station level averages, we cannot apply the instrumental variables strategy used in the previous section, which was predicated on comparing across polling stations within a precinct. Instead, the analyses in this section are based on precinct-level data. To probe causality we implement a variety of placebo, instrumental variables, and sensitivity analyses, but cannot include a precinct level fixed effect.

We begin by estimating a plain regression of trust in 2017 on recounts in 2015. The explanatory variable is the fraction of PS in a precinct that experienced recounts. We estimate the following model:

$$INE\_Impartial_s = \alpha + \beta FraccRecounted_s + X_s'\gamma + \nu_s \tag{4}$$

where $INE\_Impartial_s$ is the precinct-s average of the trust question, $FraccRecounted_s$ represents the fraction of PS presenting recounts in precinct s, and X_s is a matrix of precinct-level controls (including socioeconomic indicators from the census, average PW education, gender, and age, number of registered voters, percent vote for the three main political parties, all averaged at the precinct level). Results are shown in Table 5. The first column shows that the greater the fraction of PS with recounts in a precinct, the lower the perceived impartiality of the INE among those surveyed in that precinct. Comparing a precinct where no PS display recounts with one where all PS do, perceptions of INE impartiality are lower in the latter by -0.064, or about 13.3% of a standard deviation of the dependent variable in the regression sample.

The estimated association suggests that recounts reduce trust in INE. However, non-causal interpretations cannot be ruled out: omitted variables or reverse causality could lie behind the estimates.[28] To mitigate endogeneity concerns, we repeat the analysis substituting a "placebo" dependent variable, that is, an attitudes measure (from the same survey) that we do not expect should be influenced by past recounts. Specifically, we use the statement: "Men are better leaders and bosses than women," where, as before, respondents were asked the extent to which they agree with it (on the same five point scale as for the main dependent variable). We do not expect that recounts should affect sexist attitudes. The estimates in column 2 show there is indeed no association between recounts in 2015 and sexist attitudes in 2017. Thus, if there are omitted variables driving the precinct-level correlation between recounts and attitudes about INE's impartiality, such variables are not generating a similar association between recounts and sexist attitudes.

As a further probe of causality, we instrument recounts with inconsistencies. That is, we focus only on the variation in recounts due to inconsistencies. Concretely, we instrument the

[28]The fact that the dependent variable is measured two years after the explanatory variable to some extent argues against the possibility of reverse causality.

share of recounted PS within a precinct with the fraction of PS in the precinct that presented inconsistencies.[29] The first stage of the two-stage-least-squares estimation (i.e., the effect of inconsistencies on recounts) is powerful, with an F-stat greater than 900. The IV estimate for attitudes about INE's impartiality are reported in column 3 of Table 5. The estimate implies that trust in INE's impartiality is lower by over 20% of a standard deviation in a precinct where every PS was recounted, compared with one where no PS was recounted. This result is consistent with, and a bit stronger than, the OLS estimate in column 1. Column 4 reports the IV estimate using the placebo dependent variable, again finding no effect.

We separately estimate the sensitivity of the regression estimates to unobserved confounding. Following Oster (2017), we estimate how strong an unobserved confounder (or set of confounders) would have to be, compared with the set of included controls, to fully account for the estimated association between recounts and trust. Specifically, we compare the coefficient on recounts in column 1 of Table 5 with a similar regression with no control variables, adjusting for the changes in R^2 (details provided in the online Appendix). We find that, in order to fully account for the estimated coefficient, unobservables would be over 6 times more important than all the included controls, suggesting that omitted variable bias is unlikely to account for the estimated relationship.

As a final robustness check, we implement front-door adjustment, as developed by Pearl (1995), to estimate the causal effect of inconsistencies on trust. A key advantage of front-door adjustment is that it does not rely on the assumption that inconsistencies cause trust only via recounts, which is the exclusion restriction on which the IV estimates presented above (column 3 in Table 5) rely.[30] Front-door adjustment indicates that a precinct where all PS display some inconsistencies, compared to one where none do, is causally associated with lower trust in the electoral authorities by 3.2% of a standard deviation.[31] As an additional sanity check, we verify that inconsistencies have no effect on sexist attitudes under front-door adjustment. In sum, the battery of analyses we have presented in this section suggests that inconsistencies in vote tallies reduce trust in the electoral electoral authorities, via their effect on recounts.

[29]The exclusion restriction in this case maintains that inconsistencies at the precinct level are uncorrelated with determinants of attitudes about INE's impartiality (other than recounts), conditional on the controls. We believe this exclusion restriction is reasonably plausible, but to gain further confidence in a causal interpretation we supplement the IV evidence with a separate sensitivity analysis further below.

[30]Both the above IV analysis and the front-door analysis use three main variables: inconsistencies, recounts, and trust. One difference is that the estimands are not the same: the IV analysis estimates the effect of recounts (instrumented with inconsistencies) on trust, while the front-door adjustment estimates the "reduced-form" effect of inconsistencies on trust. A second difference is that the validity of the IV precludes the existence of common causes of inconsistencies and trust, while the validity of front-door estimates does not. See online Appendix.

[31]This is of the same order of magnitude as the equivalent coefficient from an OLS regression of trust on inconsistencies, with a full set of controls.

Table 5: Tally Inconsistencies and Trust in Electoral Institutions

	(1) INE Trust OLS	(2) Placebo OLS	(3) INE Trust IV	(4) Placebo IV
Recounts	-0.065***	-0.013	-0.138***	-0.013
	(0.02)	(0.02)	(0.04)	(0.05)
Constant	4.589***	2.898***	4.620***	2.898***
	(0.14)	(0.20)	(0.14)	(0.20)
N	6225	6225	6225	6225
R-sq	0.036	0.039	0.033	0.039
Mean of trust	4.125	4.125	4.125	3.860
Mean of PS recounted	0.528	0.528	0.528	0.528
Sd of dependent variable	0.482	0.482	0.482	0.613
F stat			981.26	981.26

This table shows the relationship between our survey measure of trust in INE in 2017 (dependent variable) and tally inconsistencies in 2015 (explanatory variable). An observation is a precinct, trust in INE and recounts are averaged at this level. Column (1) presents OLS estimates of: $INE\_Impartial_s = \alpha + \beta FraccRecounted_s + X_s'\gamma + \nu_s$, where $INE\_Impartial_s$ is the precinct-s average of the trust question, $FraccRecounted$ is the fraction of PS recounted in precinct s, and X_s is a matrix of precinct-level controls (including socioeconomic indicators from the census, average PW education, gender, and age, number of registered voters, percent vote for the three main political parties, and respondent satisfaction with democracy, all averaged at the precinct level). Column (2) displays the OLS estimates for a placebo analysis similar to column (1), but where the dependent variable is the survey item: "men are better leaders than women," which we did not expect would be affected by recounts. Columns (3) and (4) display the corresponding estimates, where we instrument recounts (explanatory variable) with presence of inconsistencies of type $j = 1, .., 3$. F-stat of the first stage is reported. *** P<.01; ** P<.05; * P<.1.

9 Conclusion

A large majority of democracies today count votes by hand. Although electronic voting has gained in popularity, growing concerns about hacking around the globe may stall its growth. However hand counting also has costs, not only in terms counting effort by citizens or salaries of workers, but also because human counting is subject to mistakes. We document that more than forty percent of polling-station level tallies display arithmetic or counting inconsistencies. We find no evidence that the inconsistencies we study are partisan in nature. But even if such inconsistencies result from honest mistakes, politicians have used them – and will continue to use them– to undermine the credibility of election results, of electoral authorities, and of democracy itself. We document that this de-legitimizing strategy may indeed work by eroding trust in the impartiality of electoral authorities, and therefore, we surmise, in democracy more generally.

Finally, we show that the education of citizens selected as poll workers, their tallying workload, and the arithmetic complexity of the tallying tasks are important causes of inconsistencies, pointing the way to policy interventions to improve the quality of vote tallies. Regarding education, our findings suggest that less-developed regions find themselves in a trap of sorts, where low levels of development feed into lower-quality vote tallies and lower trust in elections, potentially weakening accountability and the provision of good government. Regarding workload, we document that decreasing the size of the polling station will not decrease the total number of inconsistencies, as these are proportional to the number of votes.

In terms of policy, our results suggest that choices about who counts votes and what training the counters receive, may have important consequences for the quality of vote tallies. Within the set of electoral systems where votes are counted by hand, there is wide variation in terms of who tallies the vote. In many parts of Africa, PW are employees of the Electoral Commission. In New Zealand, South Korea, and some parts of the US, school-teachers do the tallying. In Sierra Leone and Zambia, PW are hired from a pool of self-selected applicants. Finally, countries such as Ecuador, Spain, and Mexico draw unpaid volunteers to function as PW. Our results on the consequences of arithmetic difficulty additionally suggest that simplifying the counting and tallying procedures might improve the accuracy of tallies, consistent with behavioral public policy guidelines (Datta and Mullainathan, 2014). Finally, our analysis suggests that the fact that voting results are imperfect, even in the absence of malfeasance, ought to receive greater scholarly attention in future work on elections, electoral behavior, and democracy.

References

Allers, M. A. and P. Kooreman (2009, Apr). More evidence of the effects of voting technology on election outcomes. *Public Choice 139*(1), 159–170.

Alvarez, R. M., M. Goodrich, T. E. Hall, D. R. Kiewiet, and S. M. Sled (2004). The complexity of the california recall election. *PS: Political Science and Politics 37*(1), 23–26.

Alvarez, R. M. and T. E. Hall (2008). *Electronic Elections: The Perils and Promises of Digital Democracy*. Princeton: Princeton University Press.

Alvarez, R. M., E. Hartman, S. Hill, and J. N. Katz (2009). Machines versus humans: The counting and recounting of pre-scored punchcard ballots.

Ansolabehere, S. and A. Reeves (2004). Using recounts to measure the accuracy of vote tabulations: Evidence from new hampshire elections 1946-2002.

Aparicio, J. (2009). Análisis estadístico de la elección presidencial de 2006:¿fraude o errores aleatorios? *Política y gobierno 16*(SPE2), 225–243.

Ascencio, S. J. and M. R. Rueda (2017). Partisan poll watchers and electoral manipulation. Unpublished manuscript, University of Rochester and Emory University.

Atkeson, L. R., Y. P. Kerevel, R. M. Alvarez, and T. E. Hall (2014). Who asks for voter identification? explaining poll-worker discretion. *The Journal of Politics 76*(4), 944–957.

Autheman, V. (2004). The resolution of disputes related to "election results": A snapshot of court practice in selected countries around the world. In *IFES Rules of Law Conference Paper Series*.

Beber, B. and A. Scacco (2012). What the numbers say: A digit-based test for election fraud. *Political Analysis 20*(2), 211–234.

Bertrand, M., S. Mullainathan, and E. Shafir (2004). A behavioral-economics view of poverty. *American Economic Review 94*(2), 419–423.

Cantú, F. (2014). Identifying irregularities in mexican local elections. *American Journal of Political Science 58*(4), 936–951.

Cantú, F. (2018). The fingerprints of fraud: Evidence from mexico's 1988 presidential election. *Working paper*.

Cantú, F. and S. Ley (2017). Poll worker recruitment: Evidence from the mexican case. *Election Law Journal: Rules, Politics, and Policy 16*(4), 495–510.

Cantú, F. and S. Ley (2017). Poll worker recruitment: Evidence from the mexican case. *Election Law Journal 16*(4), 495–510.

Crespo, J. A. (2006). Hablan las actas. *Las debilidades de la autoridad electoral mexicana. México: editorial Debate*.

Datta, S. and S. Mullainathan (2014). Behavioral design: a new approach to development policy. *Review of Income and Wealth 60*(1), 7–35.

Dee, T. S. (2007). Technology and voter intent: Evidence from the california recall election. *The Review of Economics and Statistics 89*(4), 674–683.

Domínguez, J. I. and J. A. McCann (1998). *Democratizing Mexico: Public opinion and electoral choices*. JHU Press.

Garner, P. and E. Spolaore (2005). Why chads? determinants of voting equipment use in the united states. *Public Choice 123*(3), 363–392.

Glynn, A. N. and K. Kashin (2018). Front-door versus back-door adjustment with unmeasured confounding: Bias formulas for front-door and hybrid adjustments with application to a job training program. *Journal of the American Statistical Association 113*(523), 1040–1049.

Goggin, S. N., M. D. Byrne, and J. E. Gilbert (2012). Post-election auditing: effects of procedure and ballot type on manual counting accuracy, efficiency, and auditor satisfaction and confidence. *Election Law Journal: Rules, Politics, and Policy 11*(1), 36–51.

Hall, T. E., J. Quin Monson, and K. D. Patterson (2009). The human dimension of elections: How poll workers shape public confidence in elections. *Political Research Quarterly 62*(3), 507–522.

Hyde, S. D. (2007). The observer effect in international politics: Evidence from a natural experiment. *World Politics 60*(1), 37–63.

Kerevel, Y. (2009). Election management bodies and public confidence in elections: Lessons from latin america. *Washington, DC: International Foundation for Electoral Systems*.

Kerr, N. N. (2014). Emb performance and perceptions of electoral integrity in africa. *Advancing electoral integrity*, 189–210.

Kobak, D., S. Shpilkin, and M. S. Pshenichnikov (2016). Integer percentages as electoral falisification fingerprints. *Annals of Applied Statistics 10*(1), 54–73.

Larreguy, H., J. Marshall, and P. Querubín (2016). Parties, brokers, and voter mobilization: How turnout buying depends upon the party's capacity to monitor brokers. *American Political Science Review 110*(1), 160–179.

Llewellyn, M. H., T. E. Hall, and R. M. Alvarez (2009). Electoral context and voter confidence: how the context of an election shapes voter confidence in the process.

Lott, J. R. (2009). Non-voted ballots, the cost of voting, and race. *Public Choice 138*(1), 171–197.

Mebane, W. R. (2004). The wrong man is president! overvotes in the 2000 presidential election in florida. *Perspectives on Politics 2*(3), 525–535.

Mebane, W. R. (2010). Fraud in the 2009 presidential election in iran? *Chance 23*(1), 6–15.

Molinar, J. (1991). *El Tiempo De La Legitimidad: Elecciones, Authoritarismo Y Democracia En Mexico*. Mexico: Cal Y Arena.

Myagkov, M., P. C. Ordeshook, and D. Shakin (2010). *The Forensics of Election Fraud: Russia and Ukraine*. Cambridge: Cambridge University Press.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Pliego Carrasco, F. (2007). *El mito del fraude electoral en México*. México: Editorial Pax México.

Posner, R. A. (2000). Florida 2000: a legal and statistical analysis of the election deadlock and the ensuing litigation. *The Supreme Court Review 2000*, 1–60.

Páez Benalcázar, A. (2017). Los ecuatorianos tienen derecho al recuento. *The New York Times ES*.

Schedler, A. (2009). Inconsistencias contaminantes: gobernación electoral y conflicto post-electoral en las elecciones presidenciales de 2006.

Serra, G. (2016). Vote buying with illegal resources: Manifestation of a weak rule of law in mexico. *Journal of Politics in Latin America 8*(1), 129–150.

Simpser, A. (2012). Does electoral manipulation discourage voter turnout? evidence from mexico. *The Journal of Politics 74*(3), 782–795.

Wellman, E. I., S. D. Hyde, and T. E. Hall (2018). Does fraud trump partisanship? the impact of contentious elections on voter confidence. *Journal of Elections, Public Opinion and Parties 28*(3), 330–348.

Online Appendix

Additional details on inconsistencies data

This section presents some more statistics about our data. But before we do that we want to clarify that we do not use the universe of polling stations of INE. Of the 429,765 PS installed in the three electoral years, we drop 38,493 PS located in precincts with at least one PS *extraordinaria*. These polling stations are ver different from the rest (e.g. located in remote places, places with operational problems, etc) and we wanted to be conservative. Additionally, we drop 12,725 PS located in precincts with at least one PS which did not merge with one or more of the datasets described in seccion 4.

Next we present correlations between inconsistencies in both margins: dummies indicating the presence of inconsistency type j, and an additional table of correlations of the extent of the different types of inconsistencies.

Table A1: Presence of inconsistencies correlations

	Inconsistency 1	Inconsistency 2	Inconsistency 3	Inconsistency 4
Inconsistency 1	1.00			
Inconsistency 2	0.11	1.00		
Inconsistency 3	0.06	0.17	1.00	
Inconsistency 4	0.05	0.49	0.23	1.00

This table shows the correlation of presence of inconsistencies within *acta*. Presence of inconsistency is a dummy variable indicating inconsistency of type $j = 1, .., 4$ (as defined in Section 4.1) was greater than zero.

Table A2: Extent of inconsistencies correlations

	Inconsistency 1	Inconsistency 2	Inconsistency 3	Inconsistency 4
Inconsistency 1	1.00			
Inconsistency 2	0.75	1.00		
Inconsistency 3	0.07	0.22	1.00	
Inconsistency 4	0.07	0.17	0.39	1.00

This table shows the correlation of the absolute number of inconsistencies within *acta* (as defined in Section 4.1).

Do inconsistencies reflect partisan tampering?

We first study the correlation between the party vote and the extent of inconsistencies. For each type of inconsistency $j \in \{1, 2, 3, 4\}$ and for each of the major political parties $k \in$ {PRI, PAN, PRD} we estimate the following regression:

$$PartyVotes^k_{pste} = \alpha + \beta^{kj} AbsNumInc^j_{pste} + \gamma x_{pst} + n_{st} + \epsilon_{pste} \qquad (5)$$

where $PartyVotes^k_{pste}$ denotes the number of votes for party k in polling station p within precinct s in election type e (presidential, congressional, senatorial) in election-year t, while $AbsNumInc^j_{pste}$ denotes the extent of inconsistencies of type j in absolute value as defined in Section 4.1, and n_{st} are precinct-by-year fixed effects. Including fixed effects in the regression implies that we are only making use of the variation across polling stations in a given precinct in a given year.[32] This has the advantage of controlling for all time-invariant factors that may drive inconsistencies (e.g. number of votes cast in the precinct, or average income) but it throws out all variation across precincts. Therefore we also estimate regressions without fixed effects. We estimate 24 separate regressions (3 parties $\times$ 4 inconsistency types $\times$ with vs. without FE) and report the estimates for β^{kj} in Table A3.

[32]More precisely, this is true for 2009 and 2015, while for 2012 we also make use of variation across election types (president, congress, and senate).

Table A3: Votes and inconsistencies

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	0.0070***	-0.0021	0.0020	0.00088	-0.00073	0.0022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 2	0.0062***	-0.0036***	0.0013	0.00085	-0.000071	0.0025***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 3	0.00092	-0.0022	-0.00049	-0.0026	0.0013	0.00099
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 4	-0.0024	-0.0087***	0.019***	-0.0015	-0.00060	0.0018*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PS controls	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precinct FE				Yes	Yes	Yes

This table shows the correlation between votes and inconsistencies. We estimate this (partial) correlation by estimating the following regression: $PartyVotes_{kpste} = \alpha + \beta_{kj} AbsNumInc^{j}_{pste} + n_{st} + xpst'\gamma + \epsilon_{kpste}$ where $PartyVotes_{kpste}$ counts the number of votes for party $k \in \{PRI, PAN, PRD\}$ in poll booth p of precinct s, in election year t, of election-type e (presidential, congressional, senatorial). $AbsNumInc^{j}_{pste}$ measures the inconsistencies of type $j = 1, .., 4$ (as defined in Section 4.1). For regressions corresponding to columns 1,2 and 3 we use precinct characteristics instead of precinct fixed effects n_s. We estimate 24 separate regressions (3 parties $\times$ 4 inconsistencies $\times$ with and without n_s.). These are not necessarily causal relationships, just associations. For columns 4,5 and 6 precinct FE were included. All regressions include PS controls: mean age, education and gender pf PW in each PS. Precinct controls consist of 19 sociodemographic characteristics obtained from the 2010 Census carried out by INEGI, including proportion of men, proportion of indigenous houses, employed population, average studies, and proportion of houses with certain goods, among others. Standard errors in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The estimated association between party vote and inconsistencies is substantively tiny. For ease of interpretation, we transform the estimated coefficients into the number of inconsistencies of each type j associated with 1 additional vote for party k (Table A4). For example, if an estimated coefficient were equal to .01, that would imply that 100 inconsistencies are needed to "generate" a single additional vote for the party.[33] The actual estimated correlations imply that, generally speaking, hundreds or thousands of inconsistencies would be needed to generate a single vote for any of the major political parties. For instance, 500 additional inconsistencies type 1, or 2,401 fewer inconsistencies of type 3, would be required to generate one more vote for the PRD (column 3). Given that on average there are about 8.6 type-1 inconsistencies and 2.8 type-3 inconsistencies per PS, and that the average number of voters per PS is 565 and the maximum 750, it is very difficult to argue that these inconsistencies are associated with a substantively-important misallocation of votes to parties.[34]

[33]We emphasize that these associations are not necessarily causal.

[34]The untransformed regression results are provided in Table A3 in the Appendix. Figure A1 displays an equivalent analysis at the sub-national level, with similar results.

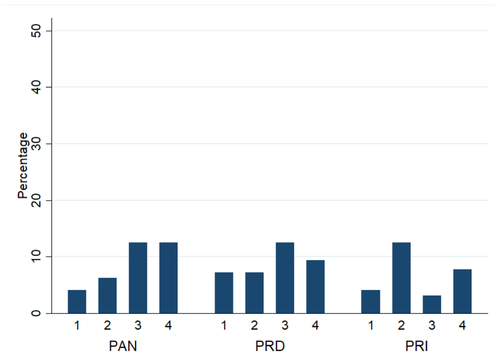
Table A4: Inconsistencies and the party vote

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	143	-476	500	1,136	-1,370	455
Inconsistency 2	161	-278	769	1,176	-14,085	400
Inconsistency 3	1,087	-455	-2,041	-385	769	1,010
Inconsistency 4	-417	-115	53	-667	-1,667	556
PS controls	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precinct-year FE				Yes	Yes	Yes

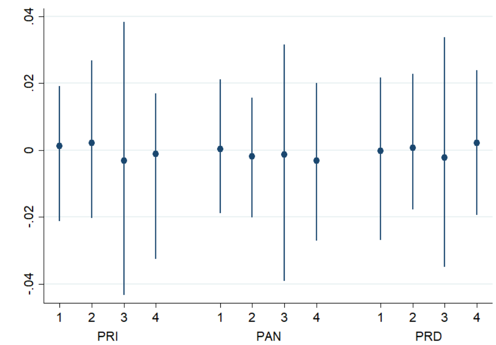
The table displays the extent of the inconsistency (i.e., the number of votes by which an accounting identity must differ) needed to add one vote to party k, on average. Columns 1-3 control for precinct characteristics, while 4-6 contain fixed effects n_{st}. We estimate 24 separate regressions (3 parties $\times$ 4 inconsistency types $\times$ with vs. without FE) and report the estimates for β^{kj} in Table A3 in the Appendix. Each cell in the current table displays the inverse of the estimated coefficients: $1/\hat{\beta^{kj}}$. All regressions include PS controls: mean age of PW at the PS, mean education of PW, and fraction male PW. Precinct controls consist of 19 different socio-demographic variables obtained from the 2010 Census including: proportion male in the precinct, proportion indigenous households, employed population, educational attainment, and proportion of households with specific goods, among others.

The previous analysis deals with national averages over all the election years in our data. It remains possible that inconsistencies could be related to party votes in particular regions and years but not in others, and that such effects could wash out in the pooled average. We therefore repeat the analysis at the state-by-election-year and district-by-election-year levels. Since there are 32 states and we have information on 3 election years, this implies that we estimate 96 coefficients for every combination of political party k and inconsistency type j, for a total of 1,152 coefficients for the state level, and 10,788 coefficients for the district level. As the table below shows, as in the national analysis, the correlation between inconsistencies and party vote is negligible: the median coefficient is zero, less than 1 percent of the coefficients are statistically significant (accounting for multiple testing), and 95 percent of the coefficients are smaller than .02 in absolute value. Figure A1 below shows the fraction of coefficients that are significant at the 5% level, as well as the distribution of their magnitudes. It turns out that 7.7% are significant at the 5% level when no multiple testing correction is done on average, but only 0.61% are significant when using a Benjamini-Holberg correction at the state level.

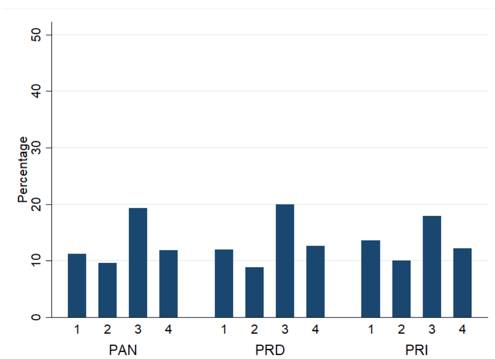
Figure A1: Inconsistencies and votes for political parties



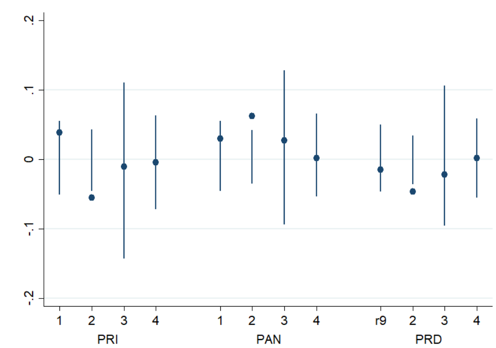
(a) Percentage of significant coefficients, state level



(b) 95% Interval of coefficients, state level



(c) Percentage of significant coefficients, district level



(d) 95% Interval of coefficients, district level

This figure shows the correlation between inconsistencies and votes for political parties. We estimate 1,152 separate regressions: one for each party, inconsistency, state and electoral year of the form $PartyVotes^k_{pste} = \alpha + \beta^{kj} NumInc^j_{pste} + x'_{pst}\gamma + \epsilon_{pste}$ where $PartyVotes^k_{pste}$ counts the number of votes for party $k \in \{PRI, PAN, PRD\}$ in poll booth p of precinct s, in election year t, of election-type e (presidential, congressional, senatorial) and x_{pst} are precinct and poll station controls. $NumInc^j_{pste}$ measures the inconsistencies of type $j = 1, .., 4$ (as defined in Section 4.1). Panel (a) shows the percentage of significant coefficients β_{kj} for each party and inconsistency. Panel (b) shows the interval [percentile 2.5, percentile 97.5] of the empiric distribution of significant coefficients for each party and inconsistency. Panel (c) and (d) are analogous to panels (a) and (b) at the district-year level.

Political party representatives and inconsistencies

We next study whether the presence of party representatives is associated with the extent of inconsistencies. We conjecture that if party representatives can exercise pressure to count or not count specific ballots then we should observe a correlation between their presence and the extent of inconsistencies. In our data, the major political parties covered a large fraction of the approximately-140,000 PS. The PRI, PAN, and PRD respectively covered 93%, 73%, and 55% of PS, on average.

For each $j \in \{1, 2, 3, 4\}$, we regress the extent of inconsistencies of type j (in absolute value) on a dummy for whether party k had representatives in polling station p within precinct s (in a given year t and election e). The model includes one such dummy for each of the major political parties:

$$AbsNumInc^{j}_{pste} = \alpha + \sum_{k} \beta^{j}_{k} RepresentativePresence_{kpste} + n_{st} + \epsilon_{pste} \quad (6)$$

Table A5 displays the results. Our preferred specification includes precinct-by-election year fixed effects (columns 5 to 8), but for completeness we also present the results without fixed effects. Most coefficients in either set of specifications are substantively small and statistically indistinguishable from zero, despite the very large number of observations.

Focusing on the specifications with fixed effects, the presence of a PAN representative, for example, is associated with an additional .21 inconsistencies of type 1 and .25 inconsistencies of type 3, but neither estimate is statistically significant (columns 5 and 7).[35] We also investigate the possibility that the presence of party representatives might moderate the relationship between inconsistencies and party votes. To test this possibility, we interact inconsistencies with an indicator for the presence of party representatives (Table A6). We find no evidence for the moderation hypothesis. In sum, the presence party representatives is not associated with the incidence of inconsistencies.

[35] Because parties choose where to send representatives, these results do not have a straightforward causal interpretation.

Table A5: Inconsistencies and party representatives

	(1) Incon. 1	(2) Incon. 2	(3) Incon. 3	(4) Incon. 4	(5) Incon. 1	(6) Incon. 2	(7) Incon. 3	(8) Incon. 4
PRI representatives	0.429	-0.228	-0.317*	-0.015	0.794	-0.391	-0.679*	0.106
	(0.34)	(0.36)	(0.17)	(0.27)	(0.83)	(0.79)	(0.36)	(0.52)
PAN representatives	-0.697***	-1.402***	0.133	0.051	0.210	-0.038	0.248	0.293
	(0.23)	(0.23)	(0.10)	(0.17)	(0.55)	(0.53)	(0.24)	(0.38)
PRD representatives	-0.168	-0.230	0.126	0.120	0.199	0.367	0.422*	0.644*
	(0.19)	(0.19)	(0.09)	(0.14)	(0.48)	(0.47)	(0.23)	(0.34)
Constant	9.433***	10.510***	3.311**	1.712	0.148	37.961	9.166	24.920
	(3.33)	(3.47)	(1.51)	(2.62)	(23.35)	(26.84)	(13.73)	(27.04)
N	437782	521736	517122	424481	437939	521915	517304	424633
R-sq	0.004	0.007	0.002	0.004	0.462	0.460	0.374	0.353
Precinct Controls	Yes	Yes	Yes	Yes				
Precinct FE					Yes	Yes	Yes	Yes

Each column corresponds to a different regression. Standard errors (clustered by precinct-year) are shown in parentheses below estimated coefficients. The unit of observation is an *acta*. Columns 1-4 include precinct-level controls, while 5-8 control for precinct fixed effects. We do not have information of inconsistency type 4 for the year 2015. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A6: Votes and party representatives interaction with inconsistencies

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	-0.005	0.004	0.006*	-0.001	-0.001	-0.001
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 2	-0.001	-0.005**	0.007**	0.000	-0.001	-0.002*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 3	-0.020**	0.004	-0.007	-0.007	0.000	-0.002
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Inconsistency 4	-0.016**	0.006	-0.008	-0.002	-0.001	-0.000
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
PS controls	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precinct FE				Yes	Yes	Yes

This table shows the correlations between the presence of political parties representatives at the PS and inconsistencies. We estimate this (partial) correlation by estimating the following regression:: $PartyVotes^k_{pste} = \alpha + \gamma^{kj} PartyPresent^k_{pst} + \delta^{kj} AbsNumInc^j_{pste} + \beta^{kj} PartyPresent^k_{pst} * AbsNumInc^j_{pste} + n_{st} + \epsilon_{pste}$. $AbsNumInc^j_{pste}$ measures the inconsistencies of type $j = 1, .., 4$, party is indexed by $k \in \{PRI, PAN, PRD\}$, poll booth by p, precinct by s, election year by $t = 2009, 2012, 2015$, and election-type by e (presidential, congressional, senatorial). The unit of observation is a PS in one election process. Columns 1 to 4 use precinct level controls, while columns 5 to 8 use precinct level fixed effects. Standard errors in parentheses are clustred at the precinct-year level. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Persistence of inconsistencies

If inconsistencies reflected partisan tampering, and if party machines varied geographically in their capacity to tamper with the vote, then one would observe that the extent of inconsistencies persists over time (and over consecutive elections) in a given geographical unit. We test for this possibility by estimating an AR(1) model for each type of inconsistency $j \in \{\text{all}, 1, 2, 3, 4\}$, where geographical unit of analysis is the precinct. We estimate the following equation:

$$\frac{AbsNumInc^j_{s,t}}{votes_{s,t}} = \alpha + \gamma^j \frac{AbsNumInc^j_{s,t-1}}{votes_{s,t-1}} + \phi_t + \nu_{s,t} \tag{7}$$

where ϕ_t are year fixed effects. The closer γ^j is to 1, the greater the persistence of inconsistency type j. For each inconsistency of type $j = 1, .., 4$, we compute the ratio of inconsistency type-j ($AbsNumInc^j$) in the precinct over the total number of votes. Table A7 presents these results. The dependent variable in column 1 is the average extent of all inconsistencies per ballot cast in precinct s in election year t. The dependent variables in columns 2-5 is the average extent of each type of inconsistencies. We find no evidence of persistence.

Table A7: Persistence of inconsistencies

	(1) All Inc.	(2) Inc. 1	(3) Inc. 2	(4) Inc. 3	(5) Inc. 4
Lag	0.001	0.002	0.001	0.000	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.072***	0.016***	0.029***	0.007***	0.020***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
N	119140	119140	119140	60477	119140
R-sq	0.001	0.000	0.000	0.000	0.000

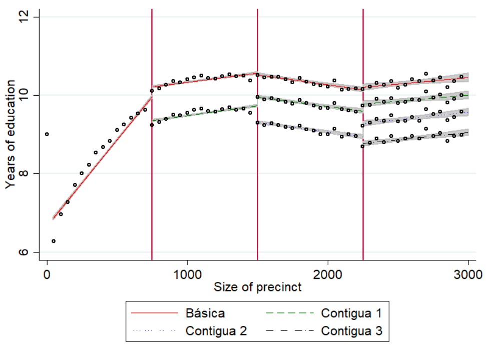
This table shows coefficients of the model: $\frac{AbsNumInc^j_{s,t}}{votes_{s,t}} = \alpha + \gamma^j \frac{AbsNumInc^j_{s,t-1}}{votes_{s,t-1}} + \phi_t + \nu_{s,t}$, where ϕ_t are year fixed effects. The closer γ^j is to 1, the greater the persistence of inconsistency type j. For each inconsistency of type $j = 1, .., 4$, we compute the ratio of inconsistency type-j ($AbsNumInc^j$) in the precinct over the total number of votes. The dependent variable in column 1 is the average extent of all inconsistencies per ballot cast in precinct s in election year t. The dependent variables in columns 2-5 is the average extent of each type of inconsistencies. We do not have data for inconsistency 4 in 2015. Standard errors shown in parentheses below coefficient estimates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

All the estimates of γ^j are very close to zero and none are statistically significant, indicating essentially no over-time persistence. This result suggests that inconsistencies have an important random component and that fixed characteristics of localities can explain only a limited amount of the variation in inconsistencies.

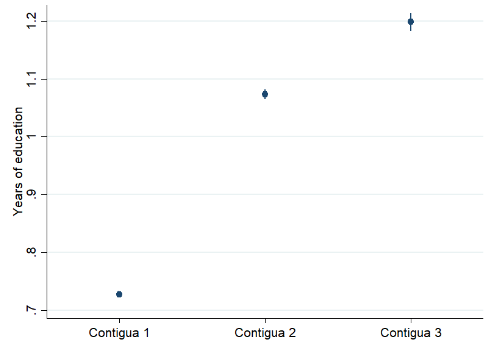
In sum, all the evidence we examined (on party votes, party representatives, and persistence), together with the scholarly consensus on the state of contemporary Mexican elections, suggest that inconsistencies do not arise out of partisan manipulation, but instead reflect primarily honest mistakes.

Causal effect of the educational attainment of PW on inconsistencies

Figure A2: Difference of years of education



(a) Years of Education by type of PS



(b) Within Precinct Differences across PS types

This figure shows the relation between average years of studies of PW designated to each PS, and type of PS (basica vs contigua). Panel (a) is a bin scatterplot (bins of size 30) of average years of education against the size of the precinct. Recall that if a precinct has between 750 and 1500 of registered voters, it must have two PS: the first is called Basica while the second is called contigua 1, and so on. The dots indicate the average education for the bin, and different line colors/patterns indicate if the average is taken for basicas, contigua 1, contigua 2, etc. It clearly shows that as we would expect from the allocation rule, PW at basicas have more education that at contigua 1, whereas contiguas themselves are ranked. Panel (b) avoids comparisons across precincts, and computes the average *difference* in education across basica and contigua 1 *within* the same precinct. It does the same comparing Basica vs contigua 1, etc. It then averages this difference across all precincts and plots this difference along with a 95% confidence interval.

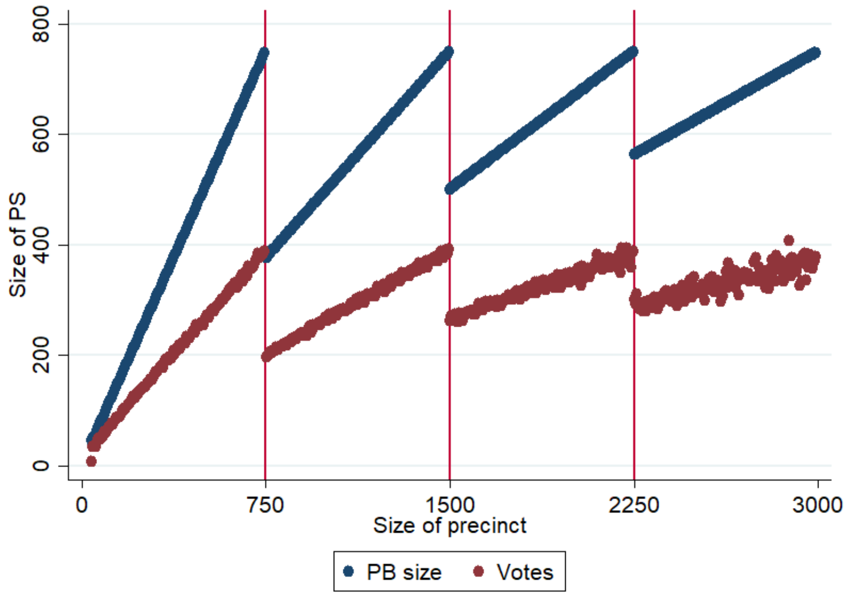
Table A8: First Stage Years of education

	(1) Years of education
Dummy PS basica	0.897***
	(0.00)
Age	-0.018***
	(0.00)
Male (%)	0.088***
	(0.00)
CAE age	0.001
	(0.00)
CAE male	0.019*
	(0.01)
CAE score	0.171***
	(0.02)
CAE education	-0.013***
	(0.00)
Constant	8.857***
	(0.15)
N	494029
R-sq	0.960
F	50967.714

This table shows the first stage of the IV estimate of inconsistencies and years of education. Taking advantage of the allocation rule of PW to PS, we instrument the average education of workers in a poll booth by an indicator of whether the poll booth is Basic or Contiguous, while controlling for precinct fixed effects and other PW average characteristics. precincts with less than 2 PS do not have contiguous polling booths and are excluded. We also exclude 994 PS which are placed in precinct with more than 14 PS. Standard errors level in parentheses. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Causal effect of workload on inconsistencies

Figure A3: **Precinct Size**



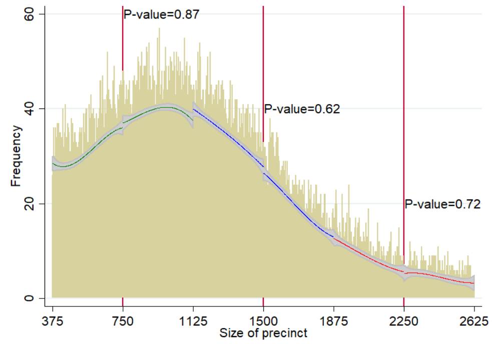
This figure shows the relation between PS size, votes cast and precinct size. Each point corresponds to the average of PS sizes and votes in a neighborhood of size 5 of precinct size. The vertical red lines correspond to precinct sizes where an additional PS is required.

Table A9: RD: Tests of Quasi-Random Assignment of Pre-Determined Characteristics

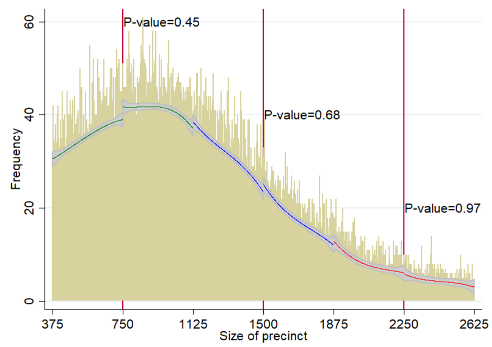
	(1) Population	(2) Houses without goods	(3) Employment Rate	(4) Male PW	(5) PW Education	(6) PW age	(7) CAE evaluation	(8) Difficulty Dummy
				Panel A: 2009 Federal Election				
Above Cutoff 751	30.127*	0.229	-0.281	0.852	0.005	-0.551	-0.007	0.023
	(18.18)	(0.28)	(0.20)	(1.26)	(0.14)	(0.35)	(0.03)	(0.03)
Above Cutoff 1501	-8.083	-0.082	-0.070	3.083**	-0.203	-0.094	0.049	-0.016
	(35.59)	(0.21)	(0.17)	(1.47)	(0.14)	(0.40)	(0.03)	(0.03)
Above Cutoff 2251	84.674	-0.414	-0.294	-0.363	-0.133	-0.000	-0.021	0.015
	(144.54)	(0.31)	(0.36)	(3.32)	(0.27)	(0.82)	(0.07)	(0.07)
N	51287	51265	51260	51281	51274	51281	47538	39452
				Panel B: 2012 Federal Election				
Above Cutoff 751	-5.952	-0.338	0.215	-1.156	0.292**	0.272	-0.002	0.041
	(15.42)	(0.26)	(0.21)	(1.19)	(0.12)	(0.34)	(0.02)	(0.03)
Above Cutoff 1501	-5.531	-0.197	-0.174	-0.459	0.209	0.603	0.015	0.016
	(30.99)	(0.24)	(0.18)	(1.48)	(0.14)	(0.42)	(0.02)	(0.03)
Above Cutoff 2251	-12.429	0.371	0.401	2.367	0.374	-1.035	-0.129***	-0.019
	(84.19)	(0.41)	(0.31)	(2.95)	(0.24)	(0.78)	(0.04)	(0.06)
N	51621	51604	51599	51618	51615	51618	48378	38340
				Panel C: 2015 Federal Election				
Above Cutoff 751	-4.332	-0.237	0.015	1.399	0.120	0.253	0.012	0.007
	(14.28)	(0.24)	(0.21)	(1.11)	(0.13)	(0.35)	(0.02)	(0.03)
Above Cutoff 1501	14.291	0.213	0.168	-1.383	-0.129	0.180	-0.048	0.026
	(23.04)	(0.24)	(0.18)	(1.41)	(0.16)	(0.43)	(0.03)	(0.03)
Above Cutoff 2251	49.495	0.181	-0.152	4.295	-0.156	-0.532	-0.003	0.036
	(71.27)	(0.36)	(0.32)	(2.75)	(0.31)	(0.81)	(0.06)	(0.07)
				Means at cutoff [-100,-1]				
Cutoff 751	911.843	2.283	95.245	41.052	11.229	37.718	8.549	0.388
Cutoff 1505	1847.132	1.684	95.237	41.395	11.785	37.208	8.582	0.348
Cutoff 2251	2988.767	1.749	95.309	39.833	11.589	36.311	8.609	0.338
N	48481	48467	48462	48481	48475	48442	46088	37011

This figure shows the balance tests for pre-treatment characteristics of the RD methodology with the precinct size as running variable. Three cutoffs of the size of precincts when PS are splitted were used: 751, 1501 and 2251. Panel A shows the RD estimation for the 2009 federal election, Panel B for the 2012 federal election and Panel C for the 2015 federal election. Panel D shows the average of the dependent variable for the interval [-100,-1] from each cutoff. For the RD a linear model with a 374 bandwidth of each cutoff was used. Clustered standard errors are reported in parenthesis. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

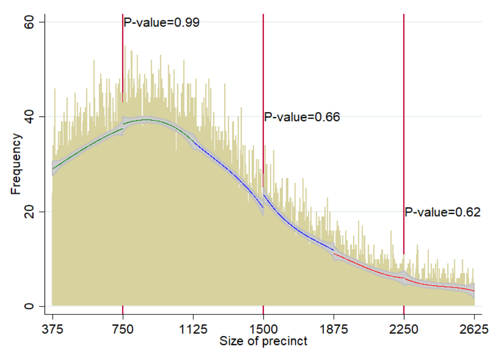
Figure A4: Distribution of size of precincts



(a) 2009 Federal Election



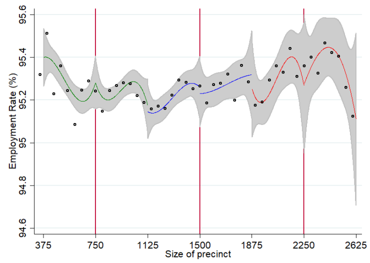
(b) 2012 Federal Election



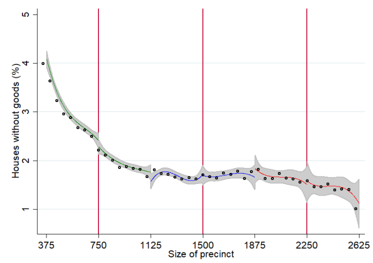
(c) 2015 Federal Election

Notes: This figure shows the distribution of precincts size for each electoral year. Lines represents a third degree polynomial approximation. The p-value of the test with null hipotesis that treatment effect is 0 is reported for each cutoff.

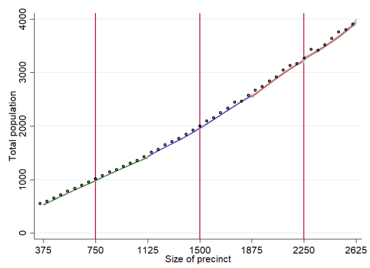
Figure A5: Pre-Treatment Characteristics



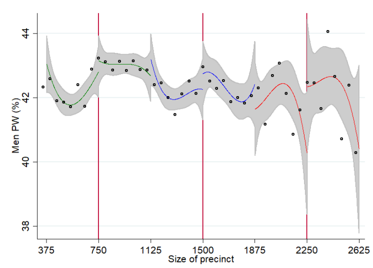
(a) Employment Rate (%)



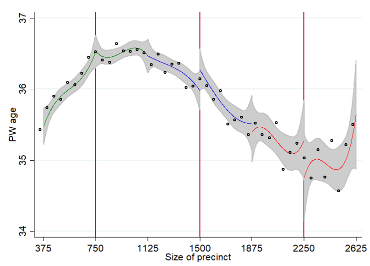
(b) Houses without goods (%)



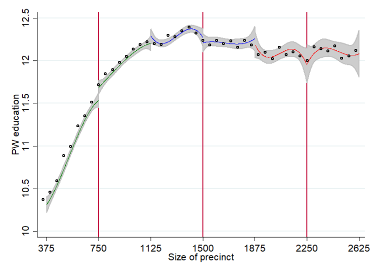
(c) Total population



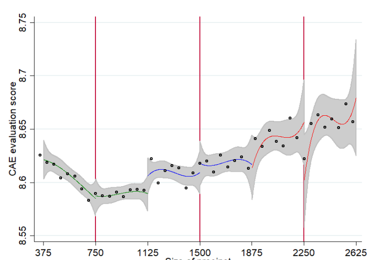
(d) % Men PW



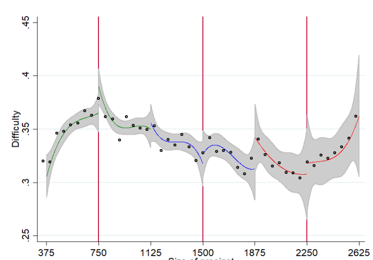
(e) PW age



(f) PW education



(g) CAE evaluation



(h) Difficulty

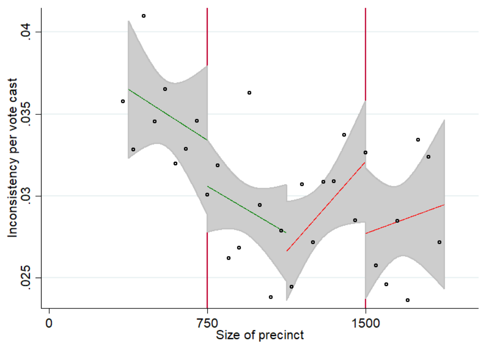
Notes: This figure shows the balance tests for pre-treatment characteristics of the RD methodology with the precinct size as running variable and split PS as treatment dummy. The unit of observation is a PS in one election process (eg. in 2012 each PS appears three times since Congressional, Senatorial and Presidential elections took place). Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections and INEGI census. Each point represents the average in a 30 precinct size neighborhood. Lines represent the RD methodology estimation and the 95% confidence interval. The RD estimate consists of a linear model with a 375 bandwitdh of each cutoff (red lines). Green line corresponds to the RD estimate of the 751 cutoff, the red line to the 1501 cutoff and the blue line to the 2251 cutoff.

Table A10: RD estimates of the effect of the PS-size-ceiling on inconsistencies

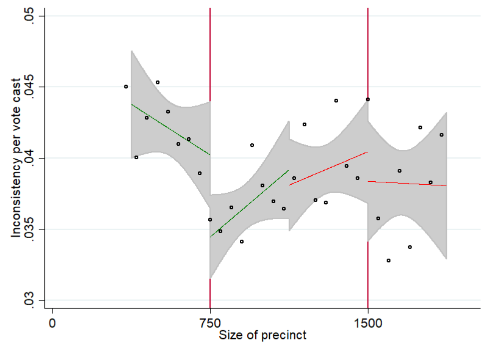
	(1) Inconsistency 1	(2) Inconsistency 2	(3) Inconsistency 3	(4) Inconsistency 4
	Panel A: RD (OLS)			
Above Cutoff 751	-5.493***	-7.607***	-1.923***	-4.014***
	(0.74)	(0.73)	(0.34)	(0.54)
Above Cutoff 1501	-4.524***	-5.449***	-1.689***	-4.194***
	(0.93)	(0.95)	(0.45)	(0.67)
Above Cutoff 2251	-3.706*	-4.504**	-1.642*	-3.652**
	(1.92)	(1.97)	(0.88)	(1.57)
	Panel B: IV			
Above Cutoff 751	0.027***	0.037***	0.009***	0.019***
	(0.00)	(0.00)	(0.00)	(0.00)
Above Cutoff 1501	0.034***	0.040***	0.012***	0.030***
	(0.01)	(0.01)	(0.00)	(0.00)
Above Cutoff 2251	0.037*	0.045**	0.017*	0.036**
	(0.02)	(0.02)	(0.01)	(0.02)
	Panel C: Means[-10,-1] from cutoff			
751	10.165	13.578	4.987	6.860
1501	13.477	17.504	5.783	13.851
2251	10.575	8.639	1.699	8.787
N	189237	225581	223440	184995
	Panel D: Testing (p-values)			
RD 751 = 1501 = 2251	0.559	0.106	0.896	0.944
RD 751 = 1501	0.415	0.072	0.678	0.835
RD 751 = 2251	0.385	0.140	0.765	0.827
RD 1501 = 2251	0.702	0.666	0.962	0.751

This table shows the effect of splitting PS on inconsistencies. Three cutoffs of the size of precincts when PS are splitted were used: 751, 1501 and 2251. Panel A shows the RD estimation with precinct size as running variable. Panel B shows the RD estimation instrumenting amount of votes. Panel C shows the average inconsistency for the interval [-100,-1] from each cutoff. Panel D reports the p-values for the test of equal coefficient between cutoffs. For the RD a linear model with a 374 bandwidth of each cutoff was used. Clustered standard errors are reported in parenthesis. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

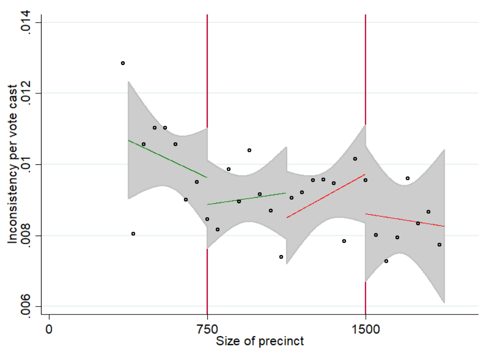
Figure A6: Effect of splitting precincts on inconsistencies per vote cast



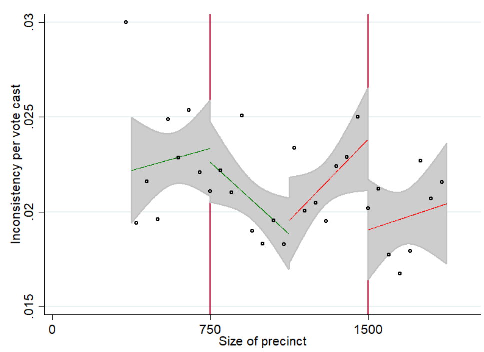
(a) Inconsistency 1



(b) Inconsistency 2



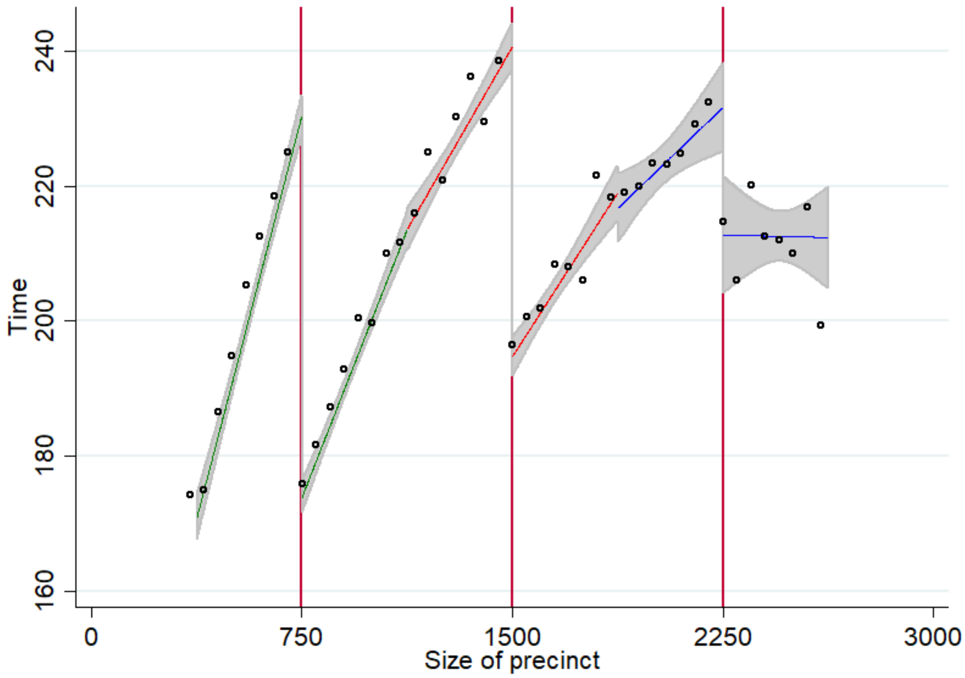
(c) Inconsistency 3



(d) Inconsistency 4

Notes: This figure shows the effect of splitting PS on inconsistencies per vote cast using a RD methodology with the precinct size as running variable and split PS as treatment dummy. The unit of observation is an *acta*. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. Each point represents the average inconsistency in a 30 precinct size neighborhood. Lines represent the RD estimate and the 95% confidence interval. The RD estimate consists of a linear model with a 374 bandwidth of each cutoff. Green line corresponds to the RD estimate of the 751 cutoff, the red line to the 1501 cutoff and the blue line to the 2251 cutoff.

Figure A7: Effect of splitting PS on counting time

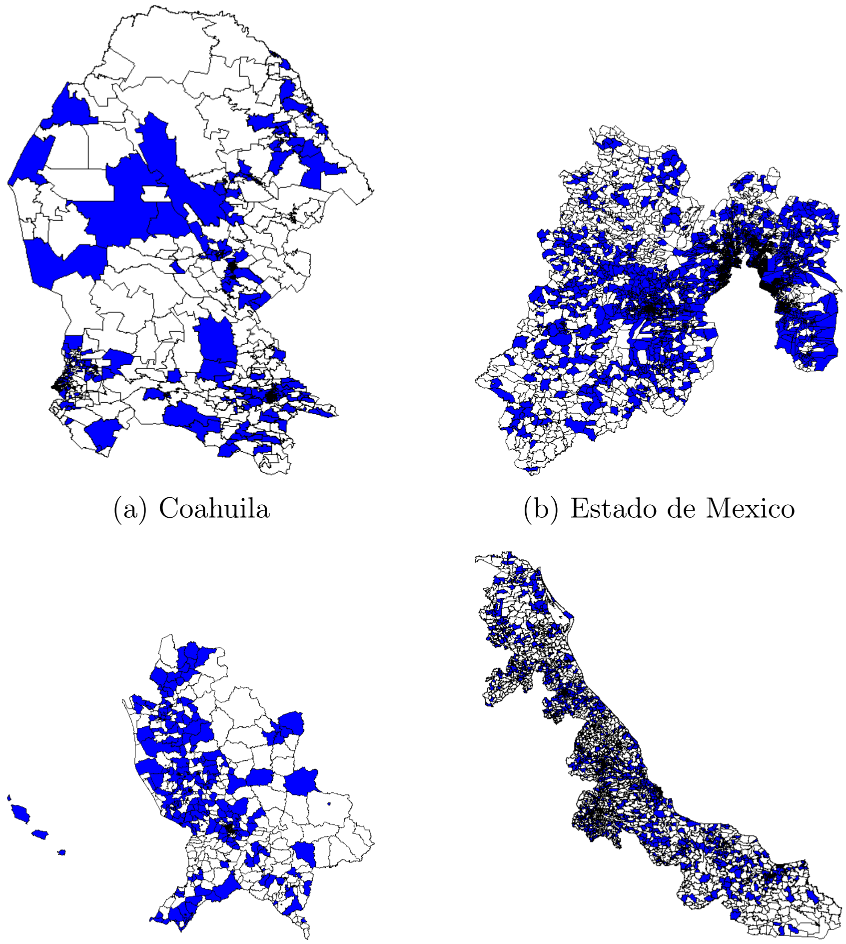


This figure shows the effect of splitting PS on time of counting using a RD methodology with the precinct size as running variable and split PS as treatment dummy. The unit of observation is an *acta*. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. Each point represents the average time in a 30 precinct size neighborhood. Lines represent the RD estimate and the 95% confidence interval. The RD estimate consists of a linear model with a 374 bandwidth of each cutoff. Green line corresponds to the RD estimate of the 751 cutoff, the red line to the 1501 cutoff and the blue line to the 2251 cutoff.

Tally quality and ballot recounts

The following Figure shows a map of recounts for 4 selected States.

Figure A8: Precincts with at least one acta recounted



(a) Coahuila (b) Estado de Mexico

(c) Nayarit (d) Veracruz

Notes: This figure shows in blue the precincts were at least one acta was recounted in 2015.

Inconsistencies as causes of recounts around the world

The following table shows that Mexico is not alone in stipulating in its electoral law that inconsistencies can trigger recounts.

Table A11: Inconsistencies and ballot recounts

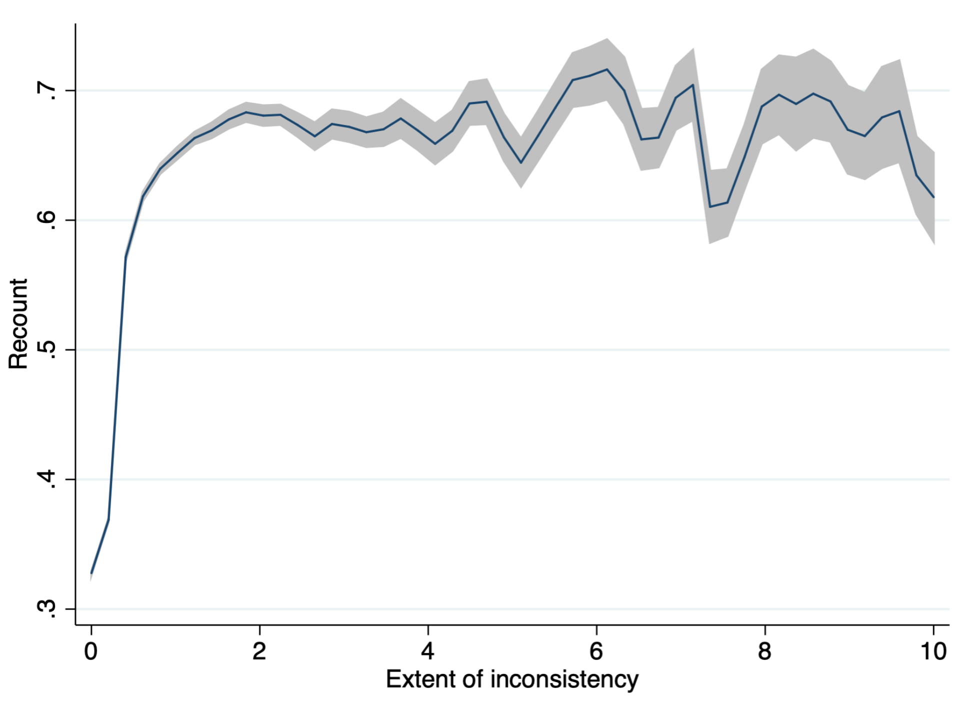
Country	**Law**
Argentina	Ley 19945, Código Nacional Electoral, art. 118: (2012)
Austria	Federal Law on Parliamentary Elections (1992), Art. 110
Brazil	"Código Eleitoral - Lei 4.737", art. 179 (II §8), 180 (II), 181: (2018); "Lei N 9.504 - das eleições", art. 88: (2018)
Chile	Electoral Law: "Ley Orgánica Constitucional sobre Votaciones Populares y Escrutinios (2016)", art. 96 & 97
Colombia	"Código Electoral", art. 122, 163, 164, 182, 189, 192: (2016)
Denmark	Parliamentary Elections and Election Administration in Denmark, Ch.5, Art. 5.2
Ecuador	"Ley Orgánica Electoral y de Organizaciones Políticas. Código de la Democracia", art. 139, 145: (2017)
Honduras	"Reglamento del Escrutinio General Definitivo en las Elecciones Generales 2009", art. 40
Mexico	"Ley General de Instituciones y Procedimientos Electorales", art. 311, 313 & 314: (2014)
Spain	Electoral Law 5/1985 of 19 June: "Ley Orgánica del Régimen Electoral General", art. 95 & 106: (2016)

This table shows a partial list of countries where inconsistent vote tallies are a legal reason for recounting ballots. A complete list of electoral laws per country can be found at https://tinyurl.com/y7b4eo23.

Relationship between inconsistencies and recounts: presence vs. extent

The following figure A9 shows a kernel regression of the relationship between the fraction of PS recounted and the extent of inconsistencies. The relationship is not linear. It seems that having a few inconsistencies (say from 1 to 20) is enough to sharply increase the likelihood of inconsistencies, with further inconsistencies beyond those not increasing the likelihood.

Figure A9: Recounts and extent of inconsistencies



This figure shows the Kernel-weighted local polynomial smoothing of the probability of recounts and average extent of inconsistencies. The average extent of inconsistencies corresponds to the average of the four types of inconsistencies at each PS. The gray area shows the 95% confidence interval.

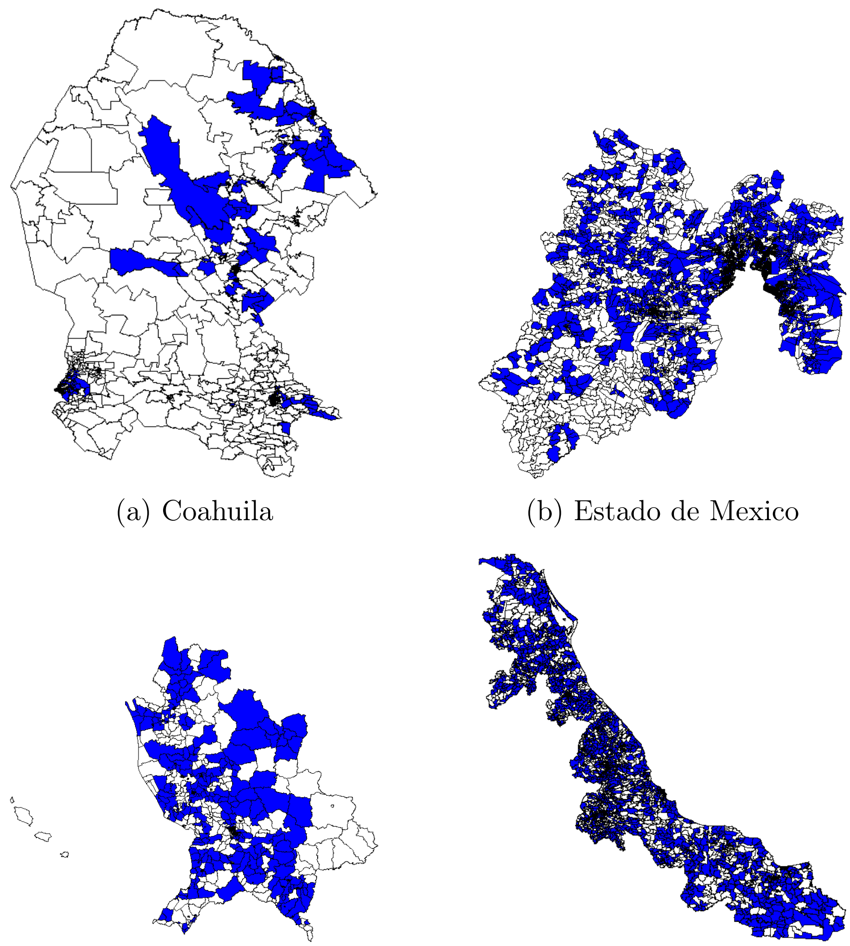
Inconsistencies and trust in the electoral authority

Poll worker survey (2017)

The survey was implemented in the 2017 local elections in Coahuila, Estado de Mexico, Nayarit and Veracruz. The survey was targeted to a random sample of almost 80,000 PW of the election during the second stage of the recruitment process. The survey was carried out by the CAE in the PW house during the first visit of the second stage. This visit corresponds to the notification of acceptance as PS with an specific role. At the moment of the survey the PW were not trained yet, so their attitudes towards INE and democracy should not have been affected.

The implementation of surveys were randomized at the CAE level. Of the 6,690 CAE, 4,014 were selected to implement the survey to all their PW. In total we have 85,006 answered surveys in 7,161 different precincts. Figure A10 shows the precincts in which at least one PW did the survey.

Figure A10: Precincts with surveys by state



(a) Coahuila (b) Estado de Mexico

(c) Nayarit (d) Veracruz

Notes: This figure shows in blue the precincts were the workers survey was implemented in each of the four states of the 2017 Federal Election.

Recounts and trust: robustness analyses

We present robustness checks on the results shown in Table Table 5 in the body of the paper. A12 and Table A13 modify the regression specification that produced in the main body of the paper in different ways, described in what follows. Table A12 controls for the extent of inconsistencies, in addition to the presence of inconsistencies. The table shows that the presence of inconsistencies continues to have considerable explanatory power, but the extent does not. This is qualitatively consistent with the pattern shown in Figure A9 in this Appendix. It is also consistent with Mexican electoral law, which specifies that it is the presence, and not the extent, of inconsistencies in an *acta* that provides grounds for a recount.

Table A12: Recounts and trust: presence vs. extent of inconsistencies

	(1) INE Trust OLS	(2) Placebo OLS	(3) INE Trust IV	(4) Placebo IV
Presence of inconsistencies	-0.043**	-0.007		
	(0.02)	(0.02)		
Extent of inconsistencies	-0.000*	0.000		
	(0.00)	(0.00)		
PS recounted			-0.121***	-0.015
			(0.04)	(0.05)
Constant	4.563***	2.892***	4.613***	2.899***
	(0.14)	(0.20)	(0.14)	(0.20)
N	6225	6225	6225	6225
R-sq	0.036	0.039	0.034	0.039
Mean of dependent variable	4.125	4.125	4.125	3.860
Mean of PS recounted	0.528	0.528	0.528	0.528
Sd of dependent variable	0.482	0.482	0.482	0.613
F stat			518.89	518.89

This table shows the relationship between "Trust in INE" and PS recounts. The difference with Table 5 in the main body of the paper is that we conlcude an additional control: the extent of inconsistencies (not just the presence). For columns (1) and (2) we estimate by OLS the following specification: $INE\_Impartial_s = \alpha + \beta FraccInconsistencies_s + \omega ExtentInconsistencies_s + X_s'\gamma + \nu_s$, where $INE\_Impartial_s$ is the precinct-s average of the aforementioned attitude question, $FraccInconsistencies_s$ represents the fraction of PS presenting inconsistencies in precinct s, $ExtentInconsistencies_s$ represents the average (absolute) number of inconsistencies in precinct s and X_s is a matrix of precinct-level controls (including socioeconomic indicators from the census, average PW education, gender, and age, number of registered voters, percent vote for the three main political parties, and respondent satisfaction with democracy, all averaged at the precinct level). For the placebo the question "men are better leaders than women" was used , since it should not be affected by recounts. For columns (3) and (4) we estimate an instrumental variable regression, were we instrument recounts with presence and extent of inconsistencies of type $j = 1, .., 3$. F stat of the first stage is reported. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A13 is also different from 5 in the main body of the paper. The first two columns weight every observation by the number of surveys completed in the corresponding precinct. This gives greater weight to precincts with more information on the attitudes variables. The last two columns repeat the analysis at the PS level (instead of at the precinct level). The explanatory variable in this case is a dummy variable indicating whether the PS was recounted (=1) or not (=0). The dependent variable continues to be the precinct-level attitudes measure (repeated for all PS within the precinct), since there are too few surveys to construct this variable at the PS level. The results continue to hold up.

Table A13: Recounts and trust: further robustness checks

	(1) INE Trust OLS	(2) Placebo OLS	(3) INE Trust IV	(4) Placebo IV
PS recounted	-0.060***	-0.138***	-0.048***	-0.078**
	(0.02)	(0.04)	(0.01)	(0.03)
Constant	4.516***	4.620***	4.448***	4.461***
	(0.15)	(0.14)	(0.18)	(0.18)
N	6225	6225	14224	14224
R-sq	0.040	0.033	0.029	0.028
Mean of dependent variable	4.125	4.125	4.111	4.111
Mean of PS recounted	0.528	0.528	0.540	0.540
Sd of dependent variable	0.482	0.482	0.476	0.476
F stat		981.26		1619.53
PS level			Yes	Yes
IV		Yes		Yes

This table shows alternative specifications for the relationship between "Trust in INE" and PS recounts. For columns (1) and (2) an observation is a precinct, and the only difference with Table 5 in the body of the paper is that precincts were weighted by the number of surveys implemented in that precinct. In columns (3) and (4) the difference is that the level of the observation is not the precinct but instead it is at the level of the polling station. That is, so the main explanatory variable is not the faction of PS recounted in a precinct, but rather a dummy for whether that PS was recounted. The dependent variable is still at the level of the precinct since we have too few surveys at the level of the PS. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Sensitivity analysis

We use Oster's aproximation under the assumption of proportional selection (Oster (2017), 8):

$$\beta^* \approx \tilde{\beta} - \delta(\beta^o - \tilde{\beta})\frac{R^{max} - \tilde{R}}{\tilde{R} - R^o} \tag{8}$$

where $\tilde{\beta}$ is the coefficient on recounts in the controlled regression (column 1 in Table 5), β^o is the coefficient on recounts in a bivariate regression of trust on recounts (=-0.097), $\tilde{R}$ and R^o respectively denote the R-squared associated with each of the regressions ($R^o = 0.0071$), and R^{max} is set to $1.3\tilde{R}$ as recommended by Oster's calibration exercise. Setting $\beta^* = 0$ and solving for δ yields the result in the text.

Front-door adjustment

Following Pearl (1995), the front-door adjusted estimate of the causal effect of inconsistencies on trust is:

$$\begin{aligned}\text{ATE} = &\sum_m (P(rec = m|inc = 1) \sum_a (E[trust|inc = a, rec = m]P(inc = a)) \\ &- \sum_m (P(rec = m|inc = 0) \sum_a (E[trust|inc = a, rec = m]P(inc = a))\end{aligned}$$

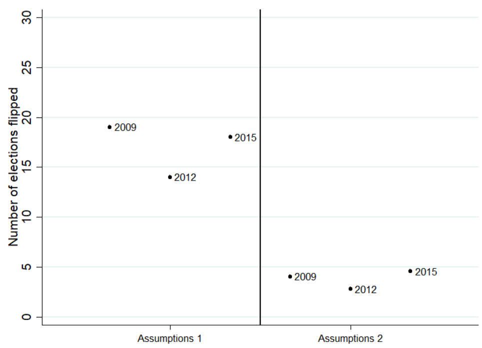
where inc stands for inconsistencies and rec for recounts. The validity of front-door adjustment relies on the alternative assumptions that: (a) the effect of inconsistencies on trust is fully mediated by recounts, and (b) recounts are unconfounded. More precisely, that the association between inconsistencies and recounts is unconfounded, and the association between recounts and trust is unconfounded after stratifying on inconsistencies. Glynn and Kashin (2018), moreover, show that the front-door adjustment can yield good results even when these assumptions do not hold exactly. Importantly, the front-door adjustment is valid even in the presence of unobserved common causes of inconsistencies and trust.

Benchmarking the consequences of inconsistencies on electoral outcomes

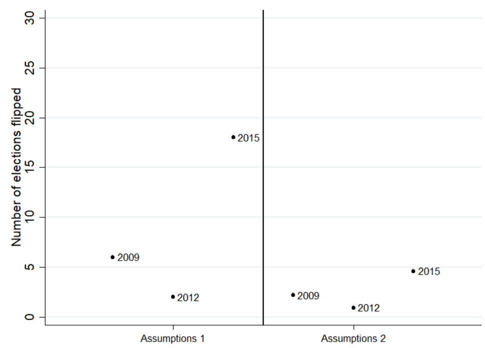
Another consequence of mistakes is that that vote counts may misrepresent who was really elected to office, especially if the margin of victory is small. Here we simulate simple back of the envelope scenarios where the the number of vote inconsistencies are assigned to parties under different assumptions. This way we explore whether the extent of tallying mistakes might conceivably have flipped elections to the national legislature. In order to implement the simulations we are forced to make some (ad-hoc) assumptions. In particular, (a) we focus on type-3 inconsistencies, since this is the only type that involves votes assigned to parties; (b) we assume that the number of inconsistencies is translated into votes one-to-one; (c) we assume that the process generating inconsistencies is independent across PS; (d) in simulation 1 we allocate the inconsistencies of a PS as votes for the first runner up; (e) in simulation 2 we allocate inconsistencies to the winner and the runner up in the PS with a 50-50 chance. In both simulation exercises we focus on congressional elections in 2009, 2012, and 2015.

In the first simulation we find that inconsistencies are prevalent enough to flip up 19 out of 300 elections in 2009. That is, 19 electoral districts in 2009 were won by a margin smaller than the total sum of inconsistencies of type 3. We view this as a worst-case scenario, since in reality it is unlikely that all the non-partisan inconsistencies, were they to be corrected, would favor just one party (the first runner up). The second simulation is a bit more realistic, and randomly assigns inconsistency-3 related votes of each PS to either the winner or the first runner up with 50-50 probability. We repeat this procedure 100 times with a different randomization for each of the 900 elections and compute the probability that each of the 900 elections is flipped. With these probabilities in hand, we estimate the expected number of flipped elections in each election year. This number is 4, 3, and 5 for 2009, 2012, and 2015 respectively. Figure A11 reports these results. We take these simulations as suggestive that, in tight races, inconsistencies could violate a basic tenet of democratic elections, namely plurality rule. But as the previous two sections show, even when inconsistencies do not flip races they can yield equally worrisome outcomes, such as spurring recounts and reducing trust in the electoral system.

Figure A11: Simulation of Inconsistencies and Elections Flipped



(a) Inconsistency 3



(b) Inconsistency 3 | Inconsistency 4 = 0

This figure shows simulations of the impact of inconsistencies in the number of elections of federal deputies election flipped, for 2009, 2012 and 2015 separately. There are two panels, both focus on type-3 inconsistency, but Panel (b) restrict the sample to Actas that had zero type-4 inconsistencies. Within each panel we present simulation 1 (left of the vertical line) and simulation 2 (right of the vertical line), which correspond to simulations under different assumptions as defined in section 9. Each dot represents the expected number that would be flipped under the simulation.