Detecting Anomalies in Data on Government Violence[*]

Kanisha D. Bond[†]
Binghamton University

Courtenay R. Conrad[‡]
University of California, Merced

Dylan Moses[§]
University of California, Merced

Joel W. Simmons[¶]
Georgetown University

Abstract

Can data on government coercion and violence be trusted when the data are generated by state itself? In this paper, we investigate the extent to which data from the California Department of Corrections and Rehabilitation (CDCR) regarding the use of force by corrections officers against prison inmates between 2008 and 2017 conform to Benford's Law. Following a growing data forensics literature, we expect misreporting of the use-of-force in California state prisons to cause the observed data to deviate from Benford's distribution. Statistical hypothesis tests and further investigation of CDCR data—which show both temporal and cross-sectional variance in conformity with Benford's Law—are consistent with misreporting of the use-of-force by the CDCR. Our results suggest that data on government coercion generated by the state should be inspected carefully before being used to test hypotheses or make policy.

Although prison officials see force as important for maintaining order inside correctional facilities (e.g., Marquardt 1986, Boin and Duin 1995), abuse has long been a point of concern among prisoners, rights advocates, and system officials. In the United States, reports of irregular documentation and excessive guard behavior have come from prisons and jails nationwide despite a host of legal and ethical mandates guiding federal, state, and local efforts to improve facility management and decrease the use of force against incarcerated citizens. Unfortunately, the accuracy, validity, and reliability of publicly-available information from within prisons—across regime type and political context—has long been questioned (e.g., Davenport 2009, Gartner and Macmillan 1995).

Despite the advantages of media-based and/or crowd-sourced data for "open-source surveillance" and chronicling police-related violence vis-á-vis the un-incarcerated public (Finch, et al. 2019), these alternatives are not always practical for understanding violence between inmates and correctional officers inside American prisons and jails. Penal institutions are among the most opaque in any political system (McDowall and Loftin 2009). With violence both among and against inmates occurring far from public view, crowdsourcing data from incarcerated citizens would be both difficult and ethically fraught, given the potential for retaliation against whistleblowers (e.g., Robbins 2016) and the political pressure to downgrade reported offenses (Maltz 2007, Seidman and Couzens 1974). Nonetheless, while the quality of prison oversight is often uncertain and accountability processes are rife with principal-agent problems (Armstrong 2014, Reiter and Koenig 2015, x-xviii), data on intra-prison violence is often singularly available from these institutions themselves.

Where the accuracy and validity of official reports may be in question and alternatives are impractical, unethical, or absent, assessing the reliability of institutionally-reported data is imperative. In this paper, we show how researchers can leverage Benford's Law, applying it to monthly count data from the California Department of Corrections (CDCR) on officers' uses of force against inmates under its jurisdiction in 37 CDCR institutions from 2008 to 2017. Specifically, we compare the distribution of the numerals that occupy the first digit of these counts to the distribution that Benford's Law expects. Our results point to persistent irregularities. We find agreement among four tests of conformity—the χ^2 goodness-of-fit test and three alternatives that are less influenced by sample size—that the data do not conform to Benford's Law. Statistical hypothesis tests and further investigation of the data show temporal and cross-sectional variance in conformity with Benford's Law and are consistent with misreporting of the use-of-force by the CDCR. Importantly,

these findings hold when accounting for the multiple comparisons problem (Benjamini and Hochberg 1995) and when using techniques informed by Bayesian analysis (Pericchi and Torres 2011; Sellke, Bayarri, and Berger 2001). Our results suggest that data on government coercion generated by the state should be inspected carefully before being used to test hypotheses or make policy.

Testing Benford's Law on CDCR Uses of Force

Benford's Law theorizes that the numerals occupying the first digit in a number—including counts of CDCR uses of force—are distributed $P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$ for all $d \in \{1, ..., 9\}$.[1] A wide array of phenomena follow the distribution—stock market prices (Ley 1996), the winning bids in auctions (Giles 2007), and the population of US counties (Hill 1995, p. 355).[2] Given widespread adherence to Benford's distribution, deviations thereof are often understood as evidence of irregular or fraudulent data (Varian 1972). The intuition behind forensics investigations is that when humans manipulate or falsify data, they are typically unable to do so in a way that adheres to the Benford distribution because of ignorance of the law's existence, psychological biases, strategic considerations, and inconsistent record keeping. Thus, violation from Benford's Law is understood as indicating human interference in the collection of data. Although legitimate data generating processes can naturally produce some deviations from Benford's Law,[3] deviations indicate—at a minimum—that data should to be inspected closely before being used for hypothesis testing.

Since 2008, the 35 institutions overseen by the California Department of Corrections have been mandated to collect and publish use-of-force data. Pursuant to California Code of Regulations Title 15, § 3268.1 (CCR 2009), "any employee who uses force or observes a staff use of force" is mandated to report details of the incident. Published monthly by the CDCR's Office of Research, each report includes counts of the use of force in each institution, disaggregated by type of force across ten categories. All uses of force can be applied against a single inmate or a group of inmates, and more than one type of force can be used per incident.[4]

We test statistical hypotheses about the extent to which CDCR data conform to Benford's Law

---

[1] Not all data are expected to conform to Benford's Law. CDCR data follow Nigrini's (2011) conformity requirements.

[2] For explanations for Benford's law, please see Miller (2015).

[3] Mebane (2013) shows that strategic voting can make unproblematic election returns data violate Benford's Law.

[4] For descriptives, please refer to our Supplemental Appendix.

and probe subsamples of our results to shed light on practical hypotheses about the extent to which CDCR data is fraudulent.[5] We use four statistical tests to evaluate the statistical conformity of CDCR use-of-force data with Benford's Law. The first is the χ^2 goodness-of-fit test, the most commonly-used test in existing scholarship. The test statistic is calculated

$$\chi^2 = \sum_{i=1}^{k} = \frac{(O_i - E_i)^2}{E_i},$$ (1)

where O_i is the observed frequency of digit i, and E_i is the frequency expected by the Benford distribution. The statistic is compared to critical values in standard χ^2 tables at 8 degrees of freedom with a 95% critical value of 15.5 and a 99% critical value of 20.1.

A drawback of the χ^2 test is that its considerable power in large samples means that even slight deviations or just a handful of outliers can result in large test statistics and correspondingly small p-values (Cho and Gaines 2007). Accordingly, our second test draws from Cho and Gaines (2007) and calculates the Euclidean distance of the observed data from the Benford distribution:

$$d = \sqrt{\sum_{i=1}^{k}(p_i - b_i)^2}.$$ (2)

In Equation (2), p_i is the proportion of observations with numeral i in the first digit, and b_i is the proportion expected by Benford's Law. As Cho and Gaines (2007) note, d is not derived from the hypothesis testing framework, so there are no critical values to compare it against. Instead, they recommend calculating $d^* = \frac{d}{1.036}$, where 1.036 is the maximum possible Euclidean distance between observed data and the Benford distribution—that is, the value that occurs when 9 is the only numeral in the first digit. Thus, d^* ranges between 0 and 1, where larger values indicate more significant deviation. For the sake of comparison, Cho and Gaines (2007, Table 1) show that 0.02 is a realistically small value of d^* that would indicate conformity and values approaching 0.1 indicate more extensive deviation. Given this, we use $d^* = 0.06$ as a useful heuristic; values below it indicate data fairly consistent with Benford's Law; values above it indicate nonconformity.

---

[5]For more on statistical and practical hypotheses, see Cleary and Thibodeau (2015, 203).

Our third test is also based on Euclidean distance. Following Morrow (2014), we calculate

$$d_N^* = \sqrt{N} * \sqrt{\sum_{i=1}^{k} (p_i - b_i)^2}. \tag{3}$$

The advantage of d_N^* over d^* is that it has critical values that allow for hypothesis testing. Specifically, Morrow (2014, pp. 4–5) shows that 1.33 is the critical value to reject the null hypothesis that the data are distributed per Benford's Law with 95% confidence, and 1.57 is sufficient to do so at the 99% confidence level.

Our fourth test is the simplest. If a set of numbers follow Benford's Law, then the numerals occupying the first digits will have a mean value of 3.441. Hicken and Mebane (2015) recommend calculating the mean from the observed data and using nonparametric bootstrapping to assess whether the 95% confidence interval contains that value.

Empirical Results

In testing whether the CDCR data conform to Benford's Law, we limit our analyses to evaluating the aggregated total uses of force, because several of the force types have counts that are too small for reliable inferences about their digit distributions. The dashed line in Figure 1 plots the proportions with which each numeral appears in the first digit of the total data, across all 35 institutions for the period from 2008 to 2017. The solid line plots the proportions expected by Benford's Law. It is evident that the observed data exhibits some inconsistencies with the expected distribution.

Aggregating CDCR data across all institutions over the entire decade conceals important nuance; further interrogation of the data is warranted to determine potential explanations for these anomalies. We rely on theoretical expectations about subsets of the data to further interrogate our results (Cleary and Thibodeau 2015, 212, Rauch et al. 2015, 262). In addition to the aggregated data being inconsistent with Benford's Law, we might also expect the data to exhibit particular temporal and cross-sectional patterns if the CDCR consistently manipulated its use-of-force data.

Consider the effect of changes in California's legal or political context on the behavior of prison guards. During the time period covered by our data, two major policy interventions were implemented that reformed when corrections officers could use force and how uses-of-force were to be reported. In

4
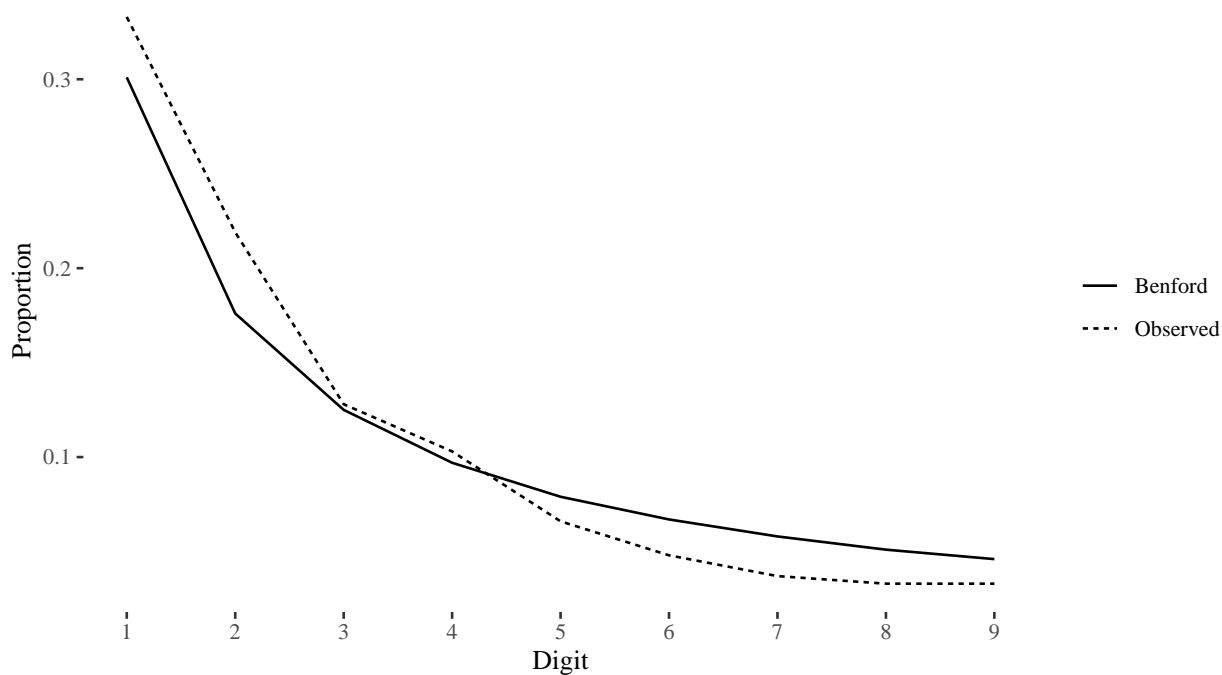
Figure 1: Observed and Expected First Digit Distributions

August 2010, the CDCR filed notice that it had adopted and implemented statewide required use-of-force reforms associated with a 1995 U.S. District Court ruling that correctional staff at Pelican Bay State Prison routinely used unusual and excessive force against inmates and were negligent in reporting it (Madrid v. Gomez, 889 F. Supp. 1146 (N.D. Cal. 1995)). A second set of policy reforms went into effect in 2016, when the state further clarified staff reporting responsibilities, updated official reporting procedures, and revised the definitions of acceptable use of force options in the CDCR Department of Operations Manual.

If these reforms had the intended effects of clarifying the conditions under which force should be used and of improving how uses of force are reported, we might expect the data to conform more closely to Benford's Law at two discrete moments. The first is in 2010, when the CDCR implemented use-of-force reforms. If these reforms worked, data from 2010 onward might adhere more closely than the data from 2008 and 2009. The second moment is in 2016, when the second set of reforms came into effect. These additional reforms might further improve the data's conformity to the Benford Distribution compared even to the 2010 reforms. Figure 2 graphs the observed and expected distributions by year and shows significant temporal variation. Compared to 2008 and
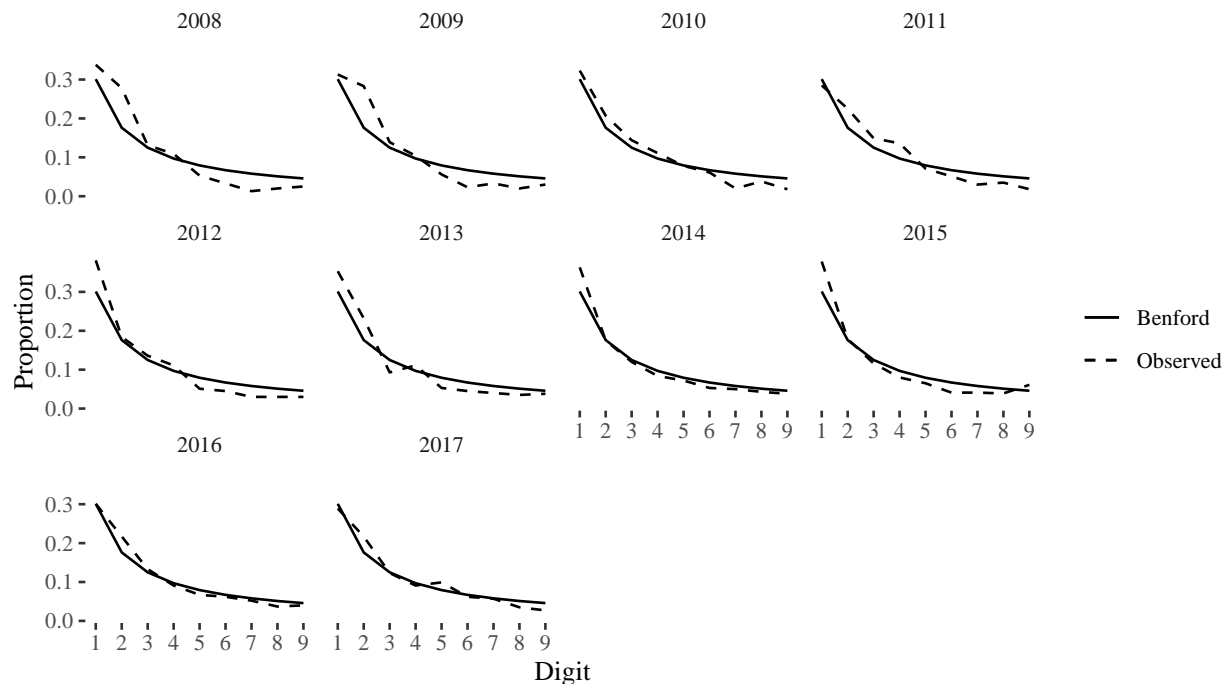
Figure 2: Digit Distributions By Year

2009, data from 2010 onward appear to conform more closely with the Benford distribution; results from 2016 and 2017 appear to conform still more closely to Benford's Law.

Table 1 shows tests of conformity to Benford's Law by year. The first row in the second column reports the χ^2 test statistic when the data are aggregated across all years and all institutions types. The test statistic is very large, allowing us to reject the null hypothesis that the observed data follow Benford's distribution. The Cho and Gaines (2007) Euclidean Distance proportion (d^*) is more conservative than the χ^2 test, but it too indicates nonconformity. Like the χ^2, Morrow's (d_N^*) Euclidean Distance test yields a large test statistic that rejects the null at the 99% level of confidence. The last two columns show similar results. If the observed data is consistent with Benford's Law, the mean of the first digit would equal 3.441. In these data, the observed mean is 2.96 with 95% confidence intervals that do not include the Benford value.

Table 1 also allows assessment of how conformity changes over time. First, though the χ^2 test statistics are very large early in the period, they decline sharply toward the end of it. According to the χ^2 test, in 2014, 2016, and 2017, we cannot reject the null hypothesis that the data appear to conform to Benford's Law. The two Euclidean distance tests and the digit mean tests also identify

6

Table 1: Tests of Conformity to Benford's Law By Year

| Year	χ^2	d^*	d_N^*	Digit Mean Test Mean	Digit Mean Test 95% CI	Obs.
All Years	159.47	0.06	4.2	3.00	2.93, 3.074	4068
2008	61.19	0.13	2.61	2.66	2.467, 2.825	396
2009	54.77	0.12	2.53	2.79	2.589, 2.991	396
2010	22.96	0.06	1.33	2.97	2.755, 3.158	396
2011	29.79	0.08	1.65	3.04	2.829, 3.233	396
2012	27.38	0.09	1.95	2.86	2.633, 3.064	396
2013	25.38	0.09	1.86	2.98	2.761, 3.203	408
2014	8.97	0.06	1.36	3.13	2.894, 3.365	420
2015	19.9	0.09	1.79	3.12	2.893, 3.363	420
2016	7.41	0.05	0.96	3.22	2.988, 3.432	420
2017	11.61	0.05	1.09	3.24	3.027, 3.459	420

Note:
Cells display test statistics in the chi^2 and d_N^* columns, proportions in the d^* column, and the first digit mean and bootstrapped confidence interval in the Digit Mean Test columns. Values in red font indicate results consistent with the data being distributed according to Benford's Law.

2016 and 2017 as years where the data deviate little from Benford's expectations. A second feature of Table 1 is the notable improvement in conformity that occurred in 2010. Compared to 2009, the χ^2 statistic in 2010 is about half the size, while the two distance tests also yield considerably smaller values. These results suggest that the 2010 reforms may have had some effect in improving how force is reported. Conformity improves even more in 2016 and 2017, which coincides with the second set of reforms implemented in 2016.

In addition to temporal variance, we investigate conformance with Benford's Law across institution type as indicative of purposeful misreporting of CDCR use-of-force data.[6] The CDCR divides its institutions into four types. New inmates arrive in the CDCR system at reception centers, where their placements needs are recorded and they are assigned to one of 34 institutions. Inmates can remain in reception centers for up to 120 days. While in reception, inmates are assigned classification scores that facilitate placement at an appropriate institution. High security institutions house the CDCR's violent male offenders; general population facilities house minimum to medium custody male inmates while providing them with opportunities for participation in vocational and academic

---

[6]Note that the frequency of CDCR institution type does not vary over time.
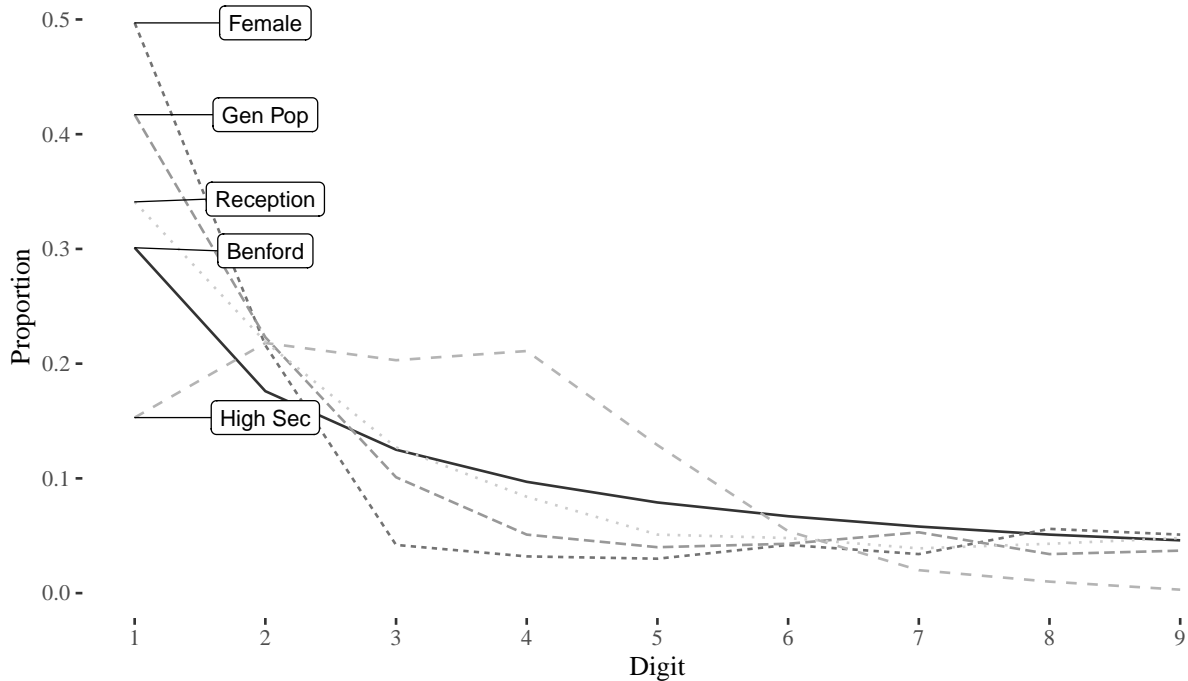
Figure 3: Digit Distribution by Institution Type

programs. Female institutions house women offenders across classification scores. Across institution types, we might expect heterogeneity in guard incentives to misreport (of fail to report) violence.

Figure 3 graphs the digit proportions for each institution type and plots them against the Benford distribution. The most glaring result is the poor fit of high security institutions, where digits 1–4 are almost uniformly distributed. General population and female institutions also fail to conform to Benford's Law; reception institutions adhere more closely to expectations. Table 2 performs the four statistical tests described above to determine whether there is heterogeneity with regard to institution type. In every case—across every test and every institution type—our tests show nonconformity with the expectations of Benford's Law. In the second column, which shows the results of the χ^2 test, and the fourth column, which shows the results of Morrow's (d_N^*) Euclidean Distance test, we are universally able to reject the null hypothesis that the observed data follow Benford's distribution. The Cho and Gaines (2007) Euclidean Distance proportion (d^*) and the test of the mean of the first digit similarly indicate reason to be concerned with potential irregularities in CDCR-reported data across institution type. Although the data do not conform to expectations for any institution type, we see the lowest conformance with Benford's Law in high security male

8

Table 2: Tests of Conformity to Benford's First Digit Distribution

| Institution Type | χ^2 | d^* | d_N^* | Digit Mean Test | | Obs. |
				Mean	95% CI	
Female	169.47	0.23	5.7	2.72	2.511, 2.935	600
Gen Pop	131.44	0.14	4.8	2.81	2.677, 2.957	1140
High Sec	446.36	0.22	7.52	3.25	3.161, 3.349	1128
Reception	48.24	0.07	2.49	3.09	2.963, 3.222	1200

Note:
Cells display test statistics in the chi^2 and d_N^* columns, proportions in the d^* column, and the first digit mean and bootstrapped confidence interval in the Digit Mean Test columns. Values in red font indicate results consistent with the data being distributed according to Benford's Law.

prisons, presumably locations where guards face high incentives to engage in violence and misreport their behavior. We see the highest conformance with Benford's Law in reception centers—locations where inmates are held for 120 days before being assigned to more permanent locations.[7]

Conclusion

In 2018, Dr. Michael Golding, chief psychiatrist for the CDCR, alleged that prison officials "played around with how they counted things and they played around with time frames" in their reports of how information regarding inmate medical and psychiatric care (Stanton 2018). This sort of "playing around" with counts and timing, if an endemic problem across CDCR institutions and across reporting areas, could account for the violations of Benford's Law we have observed. Our results suggest that there are irregularities in the CDCR's reporting of use-of-force data. The irregularities are less persistent following reforms intended to limit violence and streamline the reporting of violence, and they are the most persistent in high security male prisons. What does this variance tell us about how the likely origin of the misreporting of CDCR data?

From an organizational behavior perspective, our analyses highlight systematic differences over time and and space in how officials seem to understand the expected risks, costs, and benefits of misreporting. Prison officials—both guards and their superiors—face incentives to misreport

---

[7]In our Supplemental Appendix, we show the robustness of our results to Bayesian-influenced techniques (Pericchi and Torres 2011; Sellke, Bayarri, and Berger 2001) and accounting for the multiple comparisons problem (Benjamini and Hochberg 1995).

instances of the use of force (Maltz 2007, Seidman and Couzens 1974). Misreporting occurs for several reasons: because interactions with inmates are private, because what constitutes force is sometimes unclear, and because there is inadequate infrastructure to catch violators. Conformance with Benford's Law increased in 2010 and 2016 suggests that reforms, which made clearer what constitutes a violation and institutionalized a higher standard of reporting by increasing oversight (e.g., Cook and Fortunato 2019). That reported data gets "better" with oversight suggests purposive CDCR misreporting, although the data do not allow us to determine whether it originates from guard reports or those compiling aggregate prison data.

The contributions of this work are not limited to scholars of criminology and criminal justice. Social scientists from a variety of substantive interests are often forced to rely on state-supplied data; few explore the content of those data for signs of manipulation by the actors responsible for its generation and release. Although scholars have adduced that governments—particularly nondemocratic ones—release data on state-citizen interactions strategically (Hollyer, Rosendorff and Vreeland 2018), few have explored the content of such data for signs of manipulation. How can researchers be confident of the veracity of their data? What constitutes proper use of data pertaining to coercion that is obtained from the coercive organization itself? The forensics approach employed here can serve as a template that researchers use to more deeply and directly engage these important questions.

References

Armstrong, Andrea C. 2014. "No Prisoner Left Behind: Enhancing Public Transparency of Penal Institutions." *Stanford Law and Policy Review* 25.

Benford, Frank. 1938. "The Law of Anomalous Numbers." *Proceedings of the American Philosophical Society* pp. 551-572.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300

Boin, R. Arjen and Van Duin, Menno J. 1995. "Prison Riots as Organizational Failures: A Managerial Perspective." *The Prison Journal* 75(3):357-379.

California Codes of Regulations, Title 15, 3268.1. 1999. `https://govt.westlaw.com/calregs/Document/IC98750D4509A4142B143BD24C6DF7994?originationContext=document&transitionType=StatuteNavigator&needToInjectTerms=False&viewType=FullText&contextData=%28sc.Default%29` (September 16, 2019).

Cantú, Francisco and Sebastian M. Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's *Infamous Decade* Using Supervised Machine Learning." *Political Analysis.* 19: 409-433.

Cho, Wendy K. Tam, and Brian J. Gaines. 2007. "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance." *The American Statistician* 61(3):218-223.

Cleary, Richard J. and Jay C. Thibodeau. 2015. "Benford's Law as a Bridge between Statistics and Accounting." In *Benford's Law*, Princeton, NJ: Princeton University Press, p. 203.

Cook, Scott J. and David Fortunato. 2019. "Police (Agency) Problem: Police Use of Deadly Force." Working paper.

Davenport, Christian, 2009. *Media Bias, Perspective, and State Repression: The Black Panther Party.* Cambridge University Press.

Edwards, Frank, Michael H. Esposito, and Hedwig Lee. 2018. "Risk of Police-Involved Death by Race/Ethnicity and Place, United States, 2012-2018." *American Journal of Public Health* 108(9):1241-1248.

Finch, Brian Karl, Audrey Beck, D. Brian Burghart, Richard Johnson, David Klinger, and Kyla Thomas. 2019. "Using Crowd-Sourced Data to Explore Police-Related-Deaths in the United States (2000-2017): The Case of Fatal Encounters." *Open Health Data* 6(1):1-8.

Gartner, R., and Macmillan, R., 1995. The Effect of Victim-Offender Relationship on Reporting Crimes of Violence against Women. *Canadian Journal of Criminology* 37:393.

Giles, David E. 2007. "Benford's Law and Naturally Occurring Prices in Certain Ebay Auctions." *Applied Economics Letters* 14(3):157–161.

Hickman, Matthew J., and Stephen K. Rice. 2010. "Digital Analysis of Crime Statistics: Does Crime Conform to Benford's Law?" *Journal of Quantitative Criminology* 26(3):333-349.

Hicken, Allen and Walter R. Mebane, Jr. 2015. "A Guide to Election Forensics." Mimeograph. University of Michigan.

Hill, Theodore P. 1995. "A Statistical Derivation of the Significant-Digit Law." *Statistical Science.* 10(4): 354-363.

Hollyer, J.R., Rosendorff, B.P., & Vreeland, J.R. 2018. *Information, Democracy, and Autocracy: Economic Transparency and Political Instability.* Cambridge University Press.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mathew V. Kiang, and Justin Feldman. 2015. "Police Killings and Police Deaths are Public Health Data and Can be Counted." *PLoS Medicine* 12(12):e1001915.

Maltz, Michael. 2007. Missing UCR Data and Divergence of the NCVS and UCR Trends. In *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, ed. Addington L. Lynch J. UK: Cambridge University Press.

Marquardt, James W. 1986. "Prison Guards and the Use of Physical Coercion as a Mechanism of Prisoner Control." *Criminology* 24(2): 347-366.

Miller, Steven J. 2015. *Benford's Law.* Princeton, NJ: Princeton University Press.

noindent McDowall, David and Colin Loftin. 2009. "Do US City Crime Rates Follow a National Trend? The Influence of Nationwide Conditions on Local Crime Patterns." *Journal of Quantitative Criminology* 25(3):307-324.

Mebane, Jr, Walter. 2013. *Election Forensics.* Working Book Manuscript.

Newcomb, Simon. 1881. "Note on the Frequency of Use of the Different Digits in Natural Numbers." *American Journal of Mathematics* 4(1):39-40.

Nigrini, Mark J. 2011. *Data-Driven Forensic Investigation: Using Microsoft Access and Excel to Detect Fraud and Data Irregularities.* Hoboken: NJ: John Wiley & Sons.

Nigrini, Mark J. 2015. "Detecting Fraud and Errors Using Benford's Law." In *Benford's Law*, Princeton, NJ: Princeton University Press, p. 220.

Pericchi, Luis, and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico, and Venezuela." *Statistical Sciences* 26(4):502-516.

Robbins, Tom. 2016. "'I Was Terrified': Inmates Say They Paid a Brutal Price for a Guard's Injury," November 15. https://www.nytimes.com/2016/11/15/nyregion/new-york-prison-inmates-guards-

`beatings.html` (July 23, 2019).

Roukema, Boudewijn F. 2015. "Complementing Benford's Law for Small N: A Local Bootstrap." In *Benford's Law*, Princeton, NJ: Princeton University Press, p. 250.

Rauch, Bernhard, Max Göttsche, Gernot Brähler, and Stefan Engel. 2015. "Measuring the Quality of European Statistics." In *Benford's Law*, Princeton, NJ: Princeton University Press, p. 262.

Seidman, David and Michael Couzens. 1974. "Getting the Crime Rate Down: Political Pressure and Crime Reporting." *Law & Society Review* 8(3):457-493.

Sellke, Thomas, M. J Bayarri, and James O Berger. (2001) Calibration of ρ Values for Testing Precise Null Hypotheses. *The American Statistician* 55(1): 62-71.
Varian, H. R. 1972. Benford's Law. *American Statistician* 26(3):65.