Article

Case Studies in the Development of Reliable and Valid Social Problems Source Data

Rich Braunstein

KEYNOTE ADDRESS FOR THE 2016 GREAT PLAINS SOCIOLOGICAL ASSOCIATION ANNUAL CONFERENCE

INTRODUCTION

In his book *Damned Lies and Statistics*, Joel Best communicated that statistics are primarily social products (not social measures). Though Best focuses on the natural and mostly innocent forces that can distort data, he suggests that statistics must be approached with the skepticism of a good investigative reporter, asking questions of who created them, why were they created, what was their intended purpose, and how accurate they are (Best, 2012). The skills of thinking about data in this way are essential. Some statistics, he reports, are born bad. That is, from the start, reported statistics are sometimes based on little more than guesses or unreliable initial formations. Best's commentary on statistics that are not much good from the start illustrates the rather salient concern he raises for the adaptation or mutation of statistics that occurs downstream. Poor initial statistics and source data result from sometimes unsophisticated and at other times intentional manipulations (Best, 2012). Either way, bad statistics are powerful: They can be used to stir up public outrage or fear; they can distort our understanding of our world; and they can lead us to make poor policy choices.

Source data is of particular concern to the present work. To be clear, source data is the underlying data from which statistical analysis is conducted and public policy decisions are made. Best noted that often the validity of source data is overlooked because the underlying math appears too simple to worry about. Typically, simple-form source data is rooted in counts, averages, percentages, and rates that are included in

inferential and explanatory statistics. Best noted that we tend to take more offense to their application in advanced quantitative approaches and less offense to the shortcomings of their more simple form (Best 2012). It is at this initial level of source data that is of concern in this paper.

Examples of unreliable and/or invalid source data in public discourse and policy debate are plentiful. Without the capacity to inventory all instances of how data born bad are treated, the current paper focuses on a narrow set of concerns in the area of Native American criminal justice. This project is designed to illustrate the costs associated with bad source data and, more importantly, to consider pathways forward to overcome challenges associated with the reliance on invalid, unreliable or missing source data.

From the outset, this paper takes the position that the calculation of costs and benefits in the creation of source data has been particularly harmful to our capacity to generate reliable and valid source data. Regardless of whether this calculation is done explicitly or implicitly, the effects have been disruptive to social problems discourse. In the end, the inevitable tradeoff between the costs and benefits of acquiring good source data too often result in reduced effectiveness of social problems definitions, research, and advocacy.

Allowing for the possibility that carefully crafted cost benefit analyses can produce good source data, the high cost of acquiring valid and reliable source data commonly inhibits productivity in social problems research. This problem is exacerbated in social problems involving populations with small numbers, insofar as the relative cost of good data collection can seem greater than the potential benefits to small populations. This is clearly the case in Native American criminal justice research where limited resources inhibit the development of accurate source data from which to evaluate concerns emanating from this marginalized population. Moreover, the problem is worsened by researchers, journalists, community advocates and policy makers who accept incomplete and inaccurate measures as facts without the requisite skepticism necessary to arrive at productive social problem definitions and remedies.

The case studies from Native American criminal justice reported here involve fundamental rights and, in one instance, an issue of considerable national controversy over the past few years. The first case relates to the demographics of community policing efforts in Rapid City, South Dakota, as the Rapid City Police Department works to improve its relationship with the Native American community. The second case relates to the concern for disparate sentencing of Native Americans in the federal court system. The third case is concerned with the development and maintenance of representative jury pools in the U.S. District Court of South Dakota's Western Division where many of the federal sentences described in the second case study are determined. In sum, this paper takes up Joel Best's challenge for researchers,

community organizers and policymakers to improve standards in the identification and collection of source data.

CASE STUDIES

CASE ONE: RAPID CITY DEMOGRAPHICS

This case begins with a simple question: How many Native Americans live in Rapid City? An accurate answer to this demographic question is essential for a wide range of social problems areas involving the Native American community in Rapid City. In the way of illustration, here are sample social problem questions that rely on Native American population source data:

1) Are Native Americans subject to more traffic stops than Whites by Rapid City police officers?
2) Is intra-racial crime victimization a greater problem for Native Americans than Whites in Rapid City?
3) Are the 6th Amendment rights of Native American defendants properly supported by the jury management system in Rapid City?

The challenge of acquiring an accurate population estimate of Native American residents in the municipal context is rooted in some well-known facts. To begin, we know that the U.S. Census is subject to both over counting and under counting error (U.S. Census 2010). The understood and reported on counting error of the U.S. Census is based on the following observations, all of which apply to the Native American community in Rapid City:

- Undercounting is more likely in communities with low rates of homeownership.
- Undercounting is more likely in communities with higher rates of multigenerational households.
- Undercounting is more likely in communities with lower rates of employment.
- Undercounting is more likely in communities with fewer than 100,000 residents.
- Undercounting is more accurate among those mailing in census forms than those taken door-to-door.
- People of color and low-income people are less likely to mail in a census form.
- People of color and low-income people are less likely to be at home and accessible to door-to-door census takers.

These qualifications are largely ignored in public discussions of social problems involving the Native American community in Rapid City. A typical approach to framing the problems of Native Americans and other racial or ethnic minorities in the criminal justice system begins with a comparison of the group's population percentage in a community, as represented by the U.S. Census Bureau's most recent estimate, and the group's percentage of arrests, incarcerations, or victimizations. This much was the case in Rapid City with media reporting of a study done by the Rapid City Police Department (RCPD) by Braunstein and Schantz (2015). We all know the image of a bar chart depicting the low percentage of a minority group's population and their high rate of arrest (or other outcome of concern). In Rapid City, a leading news agency (an ABC affiliate KOTA) reported these numbers in this way on television and website broadcasts as 12% Native American population compared with 59% of Native American arrests. Alongside these disparate bars in the image were the statistics for the White community. Here it was reported that 80% of the community was White and that Whites accounted for 35% of arrests – essentially the inverse of the Native American statistics. The trouble with these population figures, and the resulting community dialogue of them, is that they are not correct.

This U.S. Census Bureau is transparent about counting error. In 2012, the U.S. Census Bureau reported "[W]hile the overall coverage of the (2010) census was exemplary, the traditional hard-to-count groups, like renters, were counted less well…Because ethnic and racial minorities disproportionately live in hard-to-count circumstances, they too were undercounted relative to the majority population." Adding to this, the U.S. Census Bureau (2012) also reported 5% undercounting of Indian Country residents. These thoughtful qualifications issued from the U.S. Census Bureau confirm there is error in the counting of urban Native Americans in Rapid City and additional error in the counting of rural Native Americans in neighboring tribal communities. Complicating the estimates, at any given time there are resident and transient populations of Native Americans in Rapid City. As such, an attempt to estimate the population for Native Americans through consideration of both the resident population from the U.S. Census Bureau (including its margin of error) and the transient population from the Department of the Interior's labor and tribal residence estimates (including its reported margin of error) can result in more precise population estimates for this unique minority racial group.

In a study contracted by RCPD, an effort was made to improve the population estimate of Native Americans living in Rapid City and to overcome the shortcomings of reporting on the single race estimates by the U.S. Census Bureau (Braunstein and Schantz 2015). The effort to revise the population of Native Americans in Rapid City began with the U.S. Census Bureau estimate of single race Native Americans in the 2010 census. This estimate was 12.4%. It continued by counting 50% of the U.S. Census Bureau estimate of multiple race individuals. This added 2.05% to the revised estimate.

Another 7.8% was added to account for the transient population of Native Americans.[1] The estimate of the population then added .62% to adjust for the U.S. Census Bureau undercount of "hard to count" Native residents in Rapid City.[2] Another .62% (or 5% undercount adjustment) was added for the historic resistance of Native Americans to participate in U.S. Federal Government surveys, effectively doubling the adjustment for the undercount of Native Americans based on (1) socioeconomic factors noted by the U.S. Census Bureau and (2) historical trust concerns of the Native American community.[3]

The sum of these percentages estimated that the Native American population of individuals living in Rapid City was 23.49% of the city's population. This revised percentage was nearly twice the U.S. Census Bureau estimate commonly reported in applied research, media presentations, and advocacy statements regarding the general welfare of Native Americans in Rapid City. Still, the RCPD report did not complete the work of estimating the actual demographics of Rapid City. Similar adjustments would have to be made for the Black, Hispanic, and Asian community members before we could have a more fully accurate estimate the percentage for the White community and for the purposes of comparison.

Even with more accurate population estimates, however, it is important to keep in mind that there are persuasive objections to the use of comparisons between a racial group's population percentage and their percentage, for instance, of arrests. The intent of this discussion of improving population source data is not to advocate for these comparisons. While this point is somewhat tangential to the current thesis, it is sufficiently important to note that a far more reliable indicator of disparities in community policing data comes through stratification – that is, looking at each racial group individually and examining percentage outcomes for arrest, victimization, citations, and other involvements. On this point, it is more productive to compare the percentage of Native Americans arrested for a specific crime to percentage of Whites arrested for that same crime to determine if a crime (or other outcome) is problematic for specific racial groups.

Regardless of what methods are used to analyze social problems data, there is little doubt that reliable population source data is essential to intentional efforts to define and remedy social problems. A primary example of this comes from Braunstein and Schantz (2015) regarding police profiling of Native American community members. We know from RCPD traffic stop data that 24.1% of traffic stops from October 2013

---

[1] This percentage was estimated at 10% of the 53,602 residents estimated to live on the three reservations bordering Rapid City, as reported by the U.S. Department of the Interior (2014).

[2] This was calculated as 5% of the total population of Native Americans, adding another .62% to the resident population total for the city.

[3] See, generally, Caldwell, J. Y., Davis, J. D., Du Bois, B., Echo-Hawk, H., Erickson, J. S., Goins, R. T., Keemer, K. 2005. "Culturally competent research with Native Americans and Alaska Natives: Findings and Recommendations of the First Symposium of the Work Group on Native American Research and Program Evaluation Methodology." *Native American and Alaska Native Mental Health Research: The Journal of the National Center*, 12(1), 1-21.

through December 2014 involved Native Americans (Braunstein and Schantz 2015). Depending on which population estimate of the Native American community in Rapid City we use, this represents either a substantial overrepresentation of Native Americans in traffic stops or a slight underrepresentation. In the first formulation, where the single-race U.S. Census Bureau statistic is used, there is an 11.7% overrepresentation, which is nearly double the population estimate of 12.4%. In the second formulation, where the revised population estimate is used, there is a .61% underrepresentation in traffic stops of Native Americans. For a community and police department at odds over racial profiling, the difference between these two disparities is substantial.

Ultimately, statistics will not resolve the conflict over police profiling. The discussion of how to address perceived or actual discrimination, however, will be very different depending on what source data is adopted and used as a benchmark for progress in the relationship between stakeholder groups. For this reason, it is essential to engage in an intentional effort to calculate the most accurate population source data possible. In the context of policing in Rapid City, this effort has evolved to include both police administrators and a representative group of community leaders. The working group that has emerged from the effort to better deliver policing services and improve the relationship of stakeholder groups has been a critical step forward for Rapid City. Taking a page from Joel Best, given that relevant statistics are social products rather than discoverable truth, perhaps it is best to leave their conceptualization and construction to the societies of scholars and practitioners who are experts in their areas. This has clearly worked in Rapid City, where careful efforts to mine the police department's data to guide discussion and policy responses have been both collaborative and successful (KEVN 2017).

CASE TWO: FEDERAL JURISDICTION IN INDIAN COUNTRY

The second case involves federal sentencing of Native American defendants. This case is helpful to describe a common liability in Native American criminal justice research and practice; namely, resource scarcity. In 2003 the U.S. Sentencing Commission took up a study of Native American criminal justice in response to concerns raised that Native American defendants are treated more harshly by the federal sentencing system than if they were prosecuted by their respective states (U.S. Sentencing Commission, 2016). As part of this study, an ad hoc advisory committee was formed, and the Commission's research staff was assigned to provide analytical support to the committee.

The effort was initiated after public hearings of the U.S. Sentencing Commission detailed the perception that jurisdictional arrangements in Native American criminal justice created structural disparities that resulted in Native Americans serving more time in federal prison for the same crimes committed by non-Natives in state courts (U.S. Sentencing Commission, 2016). On the surface, there was a question of more stringent

federal sentencing than state sentencing and higher expectations of time served in federal corrections than state corrections. Below the surface was a question of the subtler impact of the presence of federal jurisdiction over major crimes in Indian Country and of inter-state variation in jurisdictional arrangements that impact state sentencing – one of the two principal data points at issue in this case. This second question is important because of the nature and design of Public Law 280, which gave the federal government jurisdiction over major crimes committed in Indian Country. Public Law 280 created a structure whereby tribes in some states could hold concurrent jurisdiction with state government and some states where tribes would hold concurrent jurisdiction with the federal government. In some cases, a single state has variation within the state, where some tribes in the state share jurisdiction with the state government while other tribes in the same state hold jurisdiction with the federal government. A study from South Dakota, finished just before the U.S. Sentencing Commission's Ad Hoc Advisory Group on Native American Sentencing Issues was convened, reported that South Dakota state judges believed that Native Americans were sentenced to longer sentences than Whites in state court because of the presence of federal jurisdiction in the state (Braunstein and Feimer 2002). Knowing this, the U.S. Sentencing Commission was challenged to develop a research design that would control for jurisdictional variation (e.g., the impact of different criminal justice systems with full federal, partial federal, and full state jurisdiction over major crimes in Indian Country). While it can be argued that a more complete data set with structural control variables is the best way to meet the information needs of an advisory group studying the impact of federal jurisdiction, the resource-driven result was a narrower focus on pre-existing data limited to federal and state sentencing alone. As such, no data for control variables were introduced in this research, and the analysis failed to show any of the stark differences that were communicated to the U.S. Sentencing Commission at its public hearings on the subject. The federal-to-state comparisons employed simply did not question the impact of the presence of federal criminal jurisdiction in Indian Country in the United States.

The reason for this omission was communicated plainly. When prompted to develop competent source data, the U.S. Sentencing Commission's research division responded that the effort would be too expensive. The fact that Native Americans are less than 3% of the U.S. population frustrated the effort at data collection and research design stages. Here, budget and staff capacity limits of the Commission's research division clearly inhibited their study of a small population phenomenon. If the problem addressed impacted a larger population, perhaps the resources necessary to develop a more valid research design, collect the requisite source data, and complete a careful investigation could have been justified. In any case, the Commission decided to exclude contextual and control variables that were needed in their 2003 effort to properly address the impact of federal jurisdiction on Native American justice concerns.

The results have been predictable. So, little was done in the 2003 effort that another call was made in 2015 to begin the effort anew. This was largely because by 2013, as reported by the U.S. Sentencing Commission (2016), the number of Native American offenders in the federal system had increased by 27.2% over the five-year period of 2008 to 2013. Moreover, in 2013, a state with the most federal jurisdiction, South Dakota[4], had the greatest disparity for Native American defendants between federal and state sentences, and a state with the least federal jurisdiction, Oklahoma[5], had the least disparity. While it is irresponsible to conclude from this simple observation that one is causally related to the other, the absence of careful study of the impact of federal jurisdiction on Native American sentencing disparities and related structural factors has yet to be done. Supporting this view, the 2015 U.S. Sentencing Commission Tribal Issues Advisory Group (TIAG) concluded, "sentencing data currently does not exist to conduct meaningful sentencing disparity analysis" (U.S. Sentencing Commission TIAG 2016:15). This is well known in 2003. In 2017, we are still waiting on reliable source data to advance analysis in this area. While we wait, the disparate conditions of Native American criminal justice continues largely unabated in the United States, creating perhaps the only context in which a class of individuals is subjected to longer sentences and higher percentages of time served, by law, because of race-related characteristics.

CASE THREE: REPRESENTATIVE JURY POOLS IN U.S. DISTRICT COURT OF SOUTH DAKOTA'S WESTERN DIVISION

The third case of Native American criminal justice presented here involves the representative quality of a federal court's jury pool. In 1968, the Jury Selection and Service Act (JSSA) declared that it was "the policy of the United States that all litigants in federal courts entitled to trial by jury shall have the right to grand and petit juries selected at random from a fair cross section of the community" (JSSA 1968). This case, like the one before, introduces the question of what source data are necessary to test whether this constitutional requirement is met.

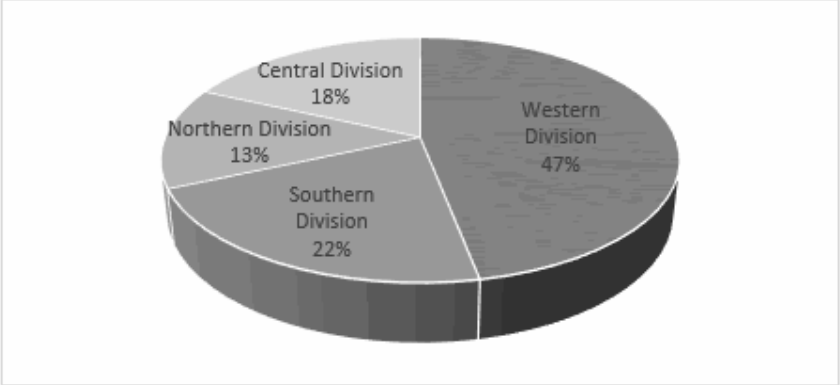
Specifically, this case presents a question of the capacity of the U.S. District Court of South Dakota's Western Division (Western Division) to maintain a representative jury pool. Keep in mind from the above discussion that 57.5% of all cases in U.S. District Court of South Dakota involve Native American defendants (USSC 2016). Additionally, the Western Division has the highest proportion of Native Americans of all four divisions of the U.S. District Court of

[4] South Dakota is one of several states to have full federal jurisdiction, meaning that 100% of Indian Country in South Dakota Tribes is subjected to the Major Crimes Act and, as a result, federal jurisdiction.

[5] Oklahoma is a unique case in that there is a large Native American population but no Indian Country within the state borders and, as a result, no federal jurisdiction over major crimes.

South Dakota, amounting to just under 50% of the entire Native American state population (see Figure 1). In summary, the federal courts in South Dakota have the highest proportion of Native American cases in the United States and most of the cases involving Native American defendants in this court come out the Western Division.

Figure One: Native Adult Population by Divisions of the US District Court of South Dakota



Source: Braunstein and Schantz 2015

We also know that while Native Americans make up approximately 24% of the Western Division's population, they make up only 6% of the division's 2013 jury pool of qualified jurors. Moreover, the percentage of Native American jurors who actually serve in criminal trials is far lower, though this is not the focus of the JSSA (only that the jury pool need be representative). A casual assessment of these disparities strongly suggest that the Western Division is not accomplishing its mandate to provide a representative jury system. A more detailed assessment, employing comparative disparity analysis typically required by courts in cross-sectional claims, demonstrated that there was a 75% difference between the Native American population's presence in the Western Division and their presence in the Western Division's qualified jury pool (Braunstein 2016). A 0% difference would mean that nearly 24% of the division's jury pool was Native American, as reflected by their estimated population presence, and a 100% difference would mean that there were no Native Americans in the jury pool. Clearly, a 75% difference between population presence and presence in the qualified jury pool is too high to serve the interest of jury section from a random and representative cross-section of the community.

These facts are well known by the United States District Court Clerk, Joseph Haas, who attributed the disparate conditions to resource limitations

associated with the Court's jury management system. According to Haas (2016), most discretionary funding available to the Court to improve the representational quality of its jury pool is dedicated to compensating and incentivizing those assigned to actual cases – that is, to get jurors to the court when they are assigned to a jury. Nevertheless, the constitutional mandate on all federal courts is to produce a representative jury pool to insure the fairness and justice of trials within the system. The essential question in the formation of representative jury pools in the federal court system is whether the distinctive group's representation in the jury pool is reasonably related to the number of the distinctive group members in the community. In less technical terms, whether there is a significant difference between the distinctive group's presence in the jury pool and the community in which they live (*Duren v. Missouri* 1979).

The development of a representative jury pool in the U.S. District Court of South Dakota's Western Division is complicated primarily by two factors very much at issue in social problems research. The first of these two factors is the distribution of Native American county population within the Western Division. The second is the lack of validity of voter registration records used as population source data by the Western Division. In a perfect world, we would like to believe that each county in the Western Division had a normal distribution of racial group residents and that voter registration is representative of Native community presence in these counties. This would make the selection of a representative jury pool simpler than it actually is. However, race in the Western Division and in South Dakota more generally, is not normally distributed. Figure 2 presents a histogram representing the distribution of Native county population in South Dakota. The graphic includes an expected curve of what a normally distributed population might look like. In other words, it presents a line under which all county populations would fall if the distribution was somewhat normal (note that it does not present expectations of a perfect bell curve given the large number of South Dakota counties with very low Native populations). The fact that more than 30 out of 66 counties have extremely low Native populations biases the distribution of state Native population downward (away from having Native county populations).

Figure 2: Native Population Histogram for All South Dakota Counties

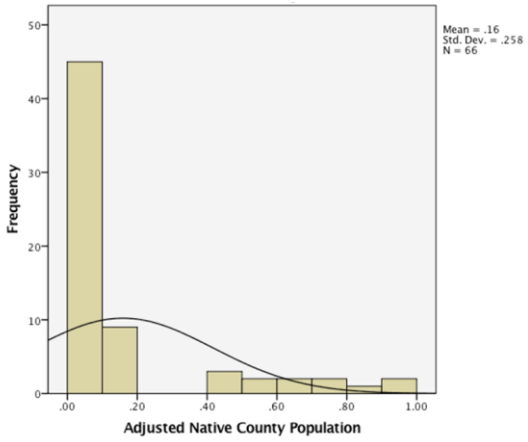
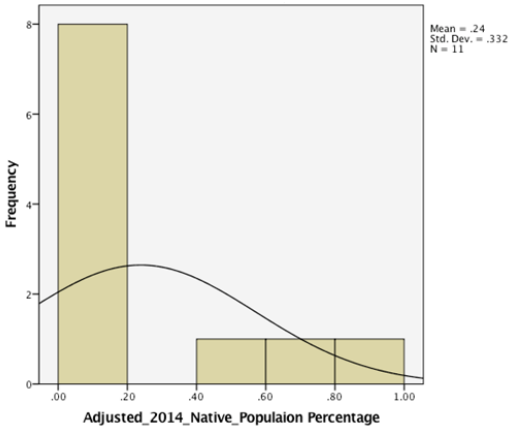


Figure 3: Western Division Counties Native Population Histogram



Source: Braunstein and Schantz 2015

A similar condition exists in the Western Division, where the distribution of race is more bifurcated. Here, counties are either a high percentage Native American or, more frequently, a very low percentage (see Figure 3). The bifurcation of Native population in the Western District makes it difficult to represent reliably Native Americans in counties with larger Native populations. Here the law of large numbers, where high percentage Native counties cancel out low percentage Native counties, does not apply because of the large number of counties with low Native populations (represented by the "spike" on the left of Figure 2). To overcome this bifurcated distribution, court administrators would need to over sample in counties with lower Native American populations and under sample in counties with higher Native American populations rather than assuming all is equal and drawing a similar number of community members from the lottery system used for the selection of the jury pool.

Variation in the response rates and resource capacities of White and Native communities in western South Dakota must also be considered when planning representative institutional structures. Reluctance of a minority group to participate can also negatively affect representative selection and, as we know from the discussion from the first case study, disproportionately impact hard to reach communities.[6] Understanding and appreciating the circumstances

[6] For discussion of United States Census Bureau under counting of 'hard to reach populations,' *see* http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html; a relevant observation includes "(w)hile the overall coverage of the census was exemplary, the traditional hard-to-count groups…were counted less well…Because ethnic and racial minorities disproportionately live in hard-to-count circumstances, they too were undercounted relative to the majority population."; *Also see* https://www.census.gov/content/dam/Census/about/about-the-bureau/Groves_Senate_Testimony_2-23-

of contemporary Native communities and their effect on behavior patterns, including complying with or responding to U.S. Government requests, requires knowledge of their history and the pain and distrust that remains today. These historically traumatic dynamics continue to affect Native people today (Caldwell, et.al. 2005) causing lower response and participation rates than non-Natives.

Through no fault of their own, federal court administrators in the Western Division begin their work structured by uneven population distributions requiring more sophisticated sampling techniques and suffering from historical distrust. Compounding these challenges is the Court's reliance on voter registration data as the single source of population data for representing the Native community. The cross-tabulation of 2012 voter registration percentages of adult county residents and Native American county population percentages showed resulted in a linear relationship between the percentage of Native Americans in a county and the percentage of citizens registered to vote where, as the percentage of Native American county population increases, voter registration decreases. South Dakota counties with low voter registration percentages tend to have moderate or high Native county population percentages. Conversely, none of the counties with high percentages of voter registration include counties with high Native populations (see Table 1).[7]

As a result of this trend, the Western Division had the lowest voter registration in 2012 and 2014 of all the U.S. District Court of South Dakota's divisions (Braunstein 2016). This was expected because the Western Division had more than twice the Native American population than any other Division (an estimated 38,125 or 46% of the entire state's Native population).

To better meet the needs and constitutional rights of the Native American community, it is necessary to supplement voter registration data with other forms of public data (e.g., driver's license, Social Security number, tribal enrollment, housing records, or some combination of these). The requirement exists because voter registration is generally not a valid proxy for population data, and its fit becomes even

10.pdf for reference to efforts that "(r)educe the undercount, especially the differential undercount which disproportionally impacts hard to count communities."

[7] In terms of the statistical significance of this relationship, the Chi-Square value was significant at the highest statistical level ($p<.01$), suggesting these observations are extremely unlikely to have resulted from chance. Similarly, the statistical correlation between these two measures is moderately strong (-.435) and statistically significant at the $p<.01$ level, reinforcing our findings from the cells of this cross-tabulation table. This analysis shows that, in South Dakota, voter registration trends are not race-neutral. The analysis shows that Native county population percentage is an effective indicator of which South Dakota counties have both high and low voter registration.

Table 1: Cross-Tabulation of County Native Population and 2012 Voter Registration Percent of Above 18 County Residents

	Low Native County Population	Moderate Native County Population	High Native County Population	Total
Low County +18 Voter Registration	3 12.50%	14 58.30%	5 38.50%	22 36.10%
Moderate County +18 Voter Registration	8 33.30%	3 12.50%	8 61.50%	19 31.10%
High County +18 Voter Registration	13 54.20%	7 29.20%	0 0.00%	20 32.80%
Percent Native Population Total	39.30%	39.30%	21.30%	100%

Source: Braunstein and Schantz 2015

worse in the context of populations with low historical voter participation rates. Its use has in federal courts has added to a substantial social problem where Native Americans, who are already disadvantaged by a disparate federal court system, are not adequately involved in the trial of their peers. Here, poor source data is compounding the negative effects of federal jurisdiction in an already disparate criminal justice context.

There are, however, reasonable fixes that the Western Division could employ without violating the Court's commitment to equal treatment for all living within the Court's jurisdiction. These include, but are not limited to, the use of supplemental data to acquire a more accurate knowledge of Native Americans living in the Western Division and a more sophisticated data management system that updates each year and does not delete confirmed data for potential jurors every two years.[8] The unfortunate reality is that, to date, insufficient resources are committed to the task. Again, we find resource limitations at the core of the problem.

CONCLUSION

The troubling realization that summarizes these cases of Native American criminal justice is that they are not instances of limitations imposed by complexity or human cognition. In the simplest terms, these are source data problems limited by resources, not possibilities. The bottom line is that the potential gains to be made from capturing and recording better quality data too often pales against the costs of ensuring

[8] Currently jury pool source data is collected every two years, at which time the previous data is expunged from the system and a follow-up response protocol for those contacted by the Court but who do not respond.

that quality. As is argued here, this is particularly the case where a small population of marginalized community members is at issue. To be sure, the path forward requires the commitment of additional public and private resources. It may be that much-needed resources will come from collaboration rather than through additional public agency expenditure, and the case of Rapid City community policing is a model for the successful collaboration that open and transparent working groups create. The paper delivered here has (hopefully) demonstrated that this additional investment is needed to support the fundamental rights of citizens, institutional priorities, and informed/productive public discourse.

In the context of Native American criminal justice, as in many other areas of social problems inquiry, we must engage in primary investigations of the relevance, timeliness, existence, coherence, completeness, and accessibility of our source data (De Veaux and Hand 2005). This is a necessary step in the 21st century, given what we know about the shortcomings of public agency budgets and the reliance of community discourse on reports and analysis of source data. The remedy is not an easy one. Often public agents and community members involved in the administration and review of source data are unaware that the datasets and findings they rely on are incomplete. This may be due to the effective use of the data for some other purpose than the task at hand and the belief that, as a result, the data is valid for secondary application. A simple reminder of the case of using voter registration data as a proxy for population data demonstrates that data that can be perfectly valid and reliable for one application (i.e. voter management) can create a host of social problems when used for another unintended, purpose (i.e. jury management). The compulsion to use pre-existing data is understandable. The alternatives typically demand more resources and engagement. In the cases noted here, alternatives involve municipalities and community leaders conducting their own population studies rather than using a nationally designed census effort that touches on the municipal level but lacks reliability in smaller population settings. The U.S. Sentencing Commission design of original research takes into account the subtle effects of federal jurisdiction rather than relying on blunt, acontextual outcome measures, and, the development of a dynamic jury management system capable of identifying, tracking, and contacting community members.

In the private sector, these challenges seem to have been overcome by the desire to generate profits and sustainable business practices. Models abound in the for-profit world that can support the effort to improve source data collection efforts for use in public policy, administration and advocacy. Examples of this include Adobe's data integration strategies to know more fully customer needs (Adobe 2016) and the analysis by the United Parcel Service (UPS 2016) of the relative cost of left turns vs. right turns by their drivers. The discussion of data mining and integration from Adobe (2016) has valuable insights for the federal jury management systems where it is important to first identify and then to stay connected to a hard to reach population through the

integration of multiple data streams. Causal analytics, as in the case of the UPS study on driving paths that reduce cost, time, and pollution have considerable research design insights for the U.S. Sentencing Commission's attempt to identify the case of growing sentencing disparities for Native American defendants. Once again, the challenges of overcoming source data that are "born bad" is not a challenge of our cumulative capacity to address social problems.

For now, the current research treats the need for enhancements in source data collection as a necessary adaptation to current practices in the social problems area. This effort corresponds with progress made to identify and implement best practices in much of what we do in the public, private, and nonprofit sectors in our 21st century society. The social problems research subfield, starting with Huff in 1954 and continued through the work of Best and others in the past decade(s) has alerted all of us to the need to overcome the problems associated with shortcuts taken in the collection and analysis of primary and secondary source data. The responsibility to do better is with all of us in the research community, public administration, the media, and in community advocacy. It begins with a healthy and much needed skepticism for the collection of source data and ends with collaboration among public and private stakeholders committed to developing a social product with greater validity and reliability than we have seen in the cases noted here and the analogous cases throughout the social problems domain.

REFERENCES

Adobe. 2016. *The Art of Integration.* https://landing.adobe.com/dam/downloads/whitepapers/261941.en.ddm-the-art-of-integration.pdf.

Best, Joel. 2012. Damned Lies and Statistics. Berkley, CA: University of California Press.

Braunstein, R. 2016. Expert Report for *U.S. v. Jason Poor Bear*, Case No. 13-CR-50156. U.S. District Court of South Dakota's Western Division.

Braunstein, R. and Feimer, S. 2002. "South Dakota Criminal Justice: A Study of Race Disparities." *South Dakota Law Review,* Volume 48.

Braunstein, R. and Schantz, T. 2015. "Rapid City Police Department and the Native Community in Rapid City: Examining Policing Trends, Community Opinions and Best Practices." https://www.rcgov.org/index.php?option=com_docman&view=download&alias=406-pd-usd-study-report&category_slug=police-department&Itemid=149.

Caldwell, J. Y., Davis, J. D., Du Bois, B., Echo-Hawk, H., Erickson, J. S., Goins, R. T., Keemer, K. 2005. "Culturally competent research with Native Americans and Alaska Natives: Findings and Recommendations of the First Symposium of the Work

Group on Native American Research and Program Evaluation Methodology." *Native American and Alaska Native Mental Health Research: The Journal of the National Center*, 12(1), 1-21.

De Veaux, Richard D. and David J. Hand. August 2005. "How to Lie with Bad Data" Statistical Science 20(3):231-238.

Duren v. Missouri. 1979. 439 U.S. 357.

Haas, J. 2016. Telephone conversation with Joseph Haas, June 30, 2016, in re *U.S. v. Jason Poor Bear*, Case No. 13-CR-50156. U.S. District Court of South Dakota's Western Division.

Jury Service and Selection Act. 1968. 28 U.S. Code § 1861.

KEVN. 2017. "Community Advisory Committee gives spring update." http://www.blackhillsfox.com/content/news/Communty-Advisory-Committee-gives-spring-updaate-422066063.html

KOTA. 2015. "RCPD Study – Native American Arrests are High." September 9, 2015.

United Parcel Service. 2016. ORION: The Algorithm Proving That Left Isn't Right. https://compass.ups.com/ups-fleet-telematics-system/

U.S. Census Bureau. 2012. http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html

U.S. Department of Interior. 2013. *American Indian Population and Labor Force Report* published by the U.S. Department of the Interior, Office of the Secretary Office of the Assistant Secretary – Indian Affairs, January 16, 2014

U.S. Sentencing Commission (USSC). 2016. "A Report of the Tribal Issues Advisory Group." https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2016/20160606_TIAG-Report.pdf