Peer Panel Review

Dr. Eric Stewart

Re: Notice of Intent to Terminate

List of Documents

1. Email from John Smith, dated 5/18/2019

2. Email from Justin Pickett, dated 6/6/2019

3. Letter to editors of *Criminology*, dated 9/24/19

4. Request for Withdrawal on 2018 article, dated 10/24/2019

5. Request for Withdrawal on 2019 article, dated 10/24/2019

6. Request for Withdrawal on 2019 and 2015 articles, dated 10/25/2019

7. Published retraction from Editors and Publisher of *Justice Quarterly*, dated 1/21/2021

8. Reanalysis Report: Johnson, Stewart, Pickett, and Gertz (2011), authored by Justin Pickett

9. Final Inquiry Report, dated 7/30/2019

10. Final Inquiry Report, dated 3/26/2020

11. Eric Stewart's Chronology Memorandum, dated 4/1/2021

12. Eric Stewart Interview, dated 4/11/22

13. Research Overview from Dean Blomberg, dated 10/20/2022

14. Talking Points from Dean Blomberg, dated 10/6/2022

15. Notice of Intent to Terminate Letter, dated 3/10/2023

Peer Panel Document
Re: Eric Stewart

#1

Hello Mr. Smith:

Thank you for bringing these concerns to our attention. We will investigate them. And if there in fact are problems, we will address them.

Eric Stewart

On May 5, 2019, at 5:55 PM, John Smith <safeguardscience@yahoo.com> wrote:

Dear Professors Johnson, Mears, and Stewart (and copied coauthors),

There seem to be irregularities in the data and findings in five articles that you published together with two surveys. This document outlines those irregularities. I am requesting the data and analysis code, or at least the R/SAS/Stata/SPSS output from your analysis, so that I can attempt to identify the source of these errors. The first survey was conducted in 2008 and the second in 2013. The five articles, grouped according to survey, follow. After acceptance and online publication, but before print publication, Mears et al. (2019, p. 487) changed all of the tables in their paper because of a "coding error." The changes removed the standard deviations (fixing the mean-SD discrepancies), added variability to the standard errors, and placed zeros in the third decimal places of coefficients and standard errors. Because the earlier "online first" version of that article is perfectly consistent with the anomalies in the other four published articles, it is discussed here to demonstrate the pattern.

2008 Survey
- Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz. (2011) Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. *Criminology, 49*(2), 401-441.
- Stewart, Eric A., Ramiro Martinez, Jr., Eric P. Baumer, and Marc Gertz. (2015) The social context of latino threat and punitive Latino sentiment. *Social Problems, 62*(1), 69-92.

2013 Survey
- Mears, Daniel P., Eric A. Stewart, Patricia Y. Warren, Miltonette O. Craig, and Ashley N. Arnio. (2019) A legacy of lynchings: Perceived black criminal threat among whites. *Law & Society Review, 53*(2), 487-517.
- Stewart, Eric A., Brian D. Johnson, Patricia Y. Warren, Jordyn L. Rosario, and Cresean Hughes. (2019) The social context of criminal threat, victim race, and punitive black and latino sentiment. *Social Problems, 66*(2), 194-221.
- Stewart, Eric A., Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio. (2018) Lynchings, racial threat, and whites' punitive views toward blacks. *Criminology, 56*(3), 455-480.

1) Anomalies in standard errors, coefficients, and p-values

In Stewart et al. (2018), tables 2-4 include regression results. The standard errors are identical to the third decimal place across models. The web links are to pictures of the tables. Stable standard errors are in yellow. Some stability in standard errors is normal, especially with a sample of this size, but this level of stability is unusual given the observed changes both in the regression coefficients and in the amount of explained variance. Combined there are 548

regression coefficients and standard errors in these three tables, but just one ends with a zero in the third decimal place. This is unusual because the distribution of third-decimal-place numbers (with rounding) should be close to uniform.

https://imgur.com/pcdB5XF

https://imgur.com/nEft7IQ

https://imgur.com/iDDs2as

The standard errors are also identical across models in Stewart et al. (2019). The following web links are to the tables. Stable standard errors are in yellow. None of the 348 regression coefficients and standard errors in these two tables end with a zero in the third decimal place, even though the distribution of third-decimal-place numbers (with rounding) should be nearly uniform.

https://imgur.com/I87pBjm

https://imgur.com/Vwbc50D

The same pattern occurred in tables 2-4 of the online-first version of Mears et al. (2019). The standard errors for the variables were identical to the third decimal place across the models. The following web links are to pictures of those tables, which were all changed between online publication and print publication. Stable standard errors are in yellow.

https://imgur.com/bJbzlh1

https://imgur.com/3fwiPr0

https://imgur.com/4IURIGr

The standard errors are also stable (to the third decimal place) across models in table 2 of Stewart et al.'s (2015) *Social Problems* article. The web link below is to the table. Stable standard errors are in yellow.

https://imgur.com/Is98UP5

The same kind of standard-error stability occurs in Johnson et al.'s (2011) table 2. The web links below are to pictures of the table, which has two panels. Stable standard errors are in yellow.

https://imgur.com/IryIT8Y

https://imgur.com/m5j3tr2

The distribution of p-values in these articles is also unusual. Across the five articles, there are 791

p-values, but not a single one falls between .045 and .105. This is highly unlikely because p-values are uniformly distributed under the null hypothesis, so one would expect numerous p-values in this range due to chance. For example, in the next six-point range, from .105 to .165, there are 35 p-values. In the six-point range after that, from .165 to .225, there are 19 p-values. And in the six-point range after that, from .225 to .285, there are 27 p-values. Actually, none of the other six-point ranges above .045 is empty, except the .045 to .105 range. This web link is to a figure that shows this.

https://imgur.com/azGDAla

2) Discordant means and standard deviations for binary variables

For binary variables, it is possible to calculate the standard deviation knowing the sample size (N) and variable mean (P), the proportion of respondents coded 1. The formula is: SD = sqrt $\{[P*(1 - P)*N] / N - 1\}$. In both Stewart et al. (2018) and the online-first version of Mears et al. (2019), the standard deviations for four of the binary variables are wildly inaccurate, too inaccurate to be due to rounding. Mears et al. (2019) changed this table before print publication. The web links are to the Stewart et al. (2018) and original Mears et al. (2019) tables with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.499), education (.497), political conservative (.495), owns home (.463).

https://imgur.com/rOzyqDF

https://imgur.com/5lS6vc

In Stewart et al. (2019), the standard deviations are wrong for nine binary variables. The web link is to the table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.50), education level (.50), political conservative (.50), owns home (.46), Southwest (.40), Northeast (.33), Midwest (.38), West (.36), and South (.50).

https://imgur.com/9dUcMUs

In Stewart et al. (2015), the standard deviations are wrong for six binary variables. The web link is to the table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.49), education level (.49), political conservative (.50), owns home (.41), Southwest (.38), and South (.50).

https://imgur.com/ShYo61j

The Johnson et al. (2011) article includes nine binary variables where the means and standard deviations do not match. The web link below is to the respective table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: white (.35), black

(.30), Hispanic (.20), married (.49), education level (.49), political conservative (.50), owns home (.41), Southwest (.38), and South (.50).

3) Identical descriptive statistics across different samples

Johnson et al. (2011:411) use a mixed-race sample in their analysis, whereas Stewart et al. (2015: 76) use a sample of non-Hispanic whites. Although these two samples have different racial compositions, they have identical means and standard deviations (to the decimal places) on twenty of the variables. It is unlikely that this is the result of the authors accidently including the wrong table in one of the papers, because the samples do differ on several variables (e.g., county ethnic and racial composition and concentrated disadvantage). The web link is to the tables. Matching descriptive statistics are in yellow. Sample racial characteristics, which are only in Johnson et al. (2011), are in red.

4) Identical descriptive statistics across samples of different sizes

Stewart et al. (2018) use a sample of 1,144 Southern whites in 90 counties. Stewart et al. (2019) use a sample of 2,408 Whites in 168 counties. Despite the large difference in sample size, these two samples have identical means and/or standard deviations on ten variables. It is unlikely that this is the result of the authors accidently including the wrong table in one of the papers, because the samples do differ on several variables (e.g., family income, political conservatism). The following web link is to the tables. Matching descriptive statistics are in yellow. Sample size differences are in red.

5) Unusual changes in sample size over time

Johnson et al. (2011) use data from a survey conducted by a polling firm called the "Research Network" in 2008. The same data are used in Stewart et al. (2015). Although the survey was conducted in 2008, the total sample size grew from N = 1,184 in Johnson et al. (2011) to N = 1,379 in Stewart et al (2015). Yet, the survey particulars remained unchanged (e.g., 54.8% response rate). And many of the descriptive statistics stayed the same (e.g., in both samples, the mean age is 47.12 and mean family income is $62,700).

6) Incorrect statistics and distributions

In the *Social Problems* article by Stewart et al. (2015, p. 76), the authors wrote: "The breakdown for annual household income was as follows: about 25 percent of the sample reported

earning less than $50,000; around 14 percent of the respondents earned between $50,000 and $75,000; 9 percent of participants earned between $75,000 and $100,000; and about 12 percent of the sample reported earning more than $100,000. The median family income in the sample is $40,900, with a mean of $62,700." There are two problems with this. First, the percentages do not add up to 100% (25% + 14% + 9% + 12% = 60%). Second, the median income cannot be $40,900 if only 25% of the sample earned less than $50,000. By definition, the median is the 50% mark.

https://imgur.com/DegoLVt

In Johnson et al. (2011, p. 412), the authors wrote: "The breakdown for annual household income was as follows: Approximately, 52.4 percent of the sample reported earning less than $50,000; approximately 21.7 percent of the respondents earned between $50,000 and $75,000; 13.9 percent of the participants earned between $75,000 and $100,000; and approximately 12.0 percent of the sample reported earning more than $100,000. The median family income in the sample is $40,900, with a mean of $62,700." If 52.4% of respondents had family incomes under $50,000 and the median family income was $40,900, then only 2.4% of respondents had a family income between $40,900 and $50,000. This is very unlikely, given the high prevalence of family incomes in this income bracket in the US population. Furthermore, given the large differences in income distributions between this paper and the 2015 *Social Problems* article (52.4% under $50K vs. 25% under $50K), it is odd that both samples have the same median ($40,900) and mean ($62,700) family incomes.

https://imgur.com/7mE8lzD

In Stewart et al. (2018, p. 471), the authors provide the only exact p-values in the study for supplementary analyses: "We thus estimated models consistent with those in tables 2 and 3 but focused instead on the Black respondents (n = 200). Black lynchings did not yield a statistically significant effect on punitive-Black sentiment (b = .065, standard error [SE] =.063, p = .35) or the Black-White punitive sentiment ratio (b = .059, SE = .058, p = .39). But the provided p-values do not match the provided coefficients and standard errors. The values should instead be: p = .30 and p = .31.

https://imgur.com/X65TQkQ

7) Unlikely survey design and data structure

In Johnson et al. (2011, p. 411-413) and Stewart et al. (2015, p.76) the authors wrote that the Research Network administered the 2008 survey using "a two-stage modified Mitofsky–Waksberg sampling design" to randomly sample "American households with either landlines or cellular Phones." This yielded a high degree of clustering, with more than ten respondents in each

county, on average. But Mitofsky–Waksberg sampling is rarely used for cell phones. Additionally, in neither study do the authors discuss how they handled errors in matching wireless numbers to counties. According to a Pew Report by Christian et al. (2009): "The geographic information derived from cell phone numbers is subject to a great deal of error ... the sample and zip code-derived county do not match for nearly four-in-ten cell respondents (39%)."

- http://www.pewresearch.org/2009/07/09/accurately-locating-where-wireless-respondents-live-requires-more-than-a-phone-number/

The 2013 survey also used "a two-stage modified Mitofsky–Waksberg sampling design" to randomly sample "households with landlines or cell phones" (Stewart et al., 2018, p. 460). Although 2013 and 2008 surveys have identical designs, both unusual, it does not appear that the Research Network administered the 2013 survey. The Research Network is not mentioned in any of the three articles that use the 2013 data.

The differences between the 2013 sample characteristics and the US population are also difficult to reconcile with the use of random sampling for a sample this large. The total sample includes 2,736 Americans, of whom 2,408 are "non-Latino White respondents (N = 2,408)" (Stewart et al., 2018, p. 461). This means 88% of respondents are non-Latino Whites. According to the US Census, however, only 60.7% of Americans are non-Latino Whites. A random sample of this size should not be 27 percentage points off the population value. Stewart et al. (2018, p. 461) write that the "southern sample [of Whites] in our analysis consists of 1,441 respondents who resided in 90 counties across these 11 states." This means that 60% of all whites in the sample (1,441/2,408) lived in 11 southern states. This is odd, because in their *Social Problems* article, Stewart et al. (2019) give a different percentage:

https://imgur.com/tP8EWhk

It is also improbable because, according to the US Census, only 37.5% of all Americans lived in the South in 2013. Moreover, the percentage of White Americans living in the South was even lower, because the South is the second most racially diverse region. In a large nationally representative survey, one would expect about 30-35% of White respondents to live in the South, not 55-60%. A large random sample should not be off by over 20 percentage points.

- https://www.census.gov/popclock/print.php?component=growth&image=//www.census.gov/popclock/share/images/growth_1530403200.png
- https://statisticalatlas.com/United-States/Race-and-Ethnicity#figure/region

None of the articles using the 2013 survey list a funding agency or grant number, which is surprising, because a nationally representative, dual-frame, telephone survey of 2,736 Americans would cost well over $100,000.

- https://www.surveypractice.org/article/3168-the-changing-costs-of-random-digital-dial-cell-phone-and-landline-interviewing

Respectfully,
John Smith

Peer Panel Document

Re: Eric Stewart

#2

From:	Pickett, Justin
To:	Diana Key
Subject:	Fw: Files and Concerns
Date:	Monday, June 10, 2019 8:40:30 AM
Attachments:	justin_voting_data.xls
	Voting - Data.xls
	Case_comparison.xls
	2009 Ethnic Threat Paper Final.ppt
	Ethnic Threat Paper 10-10-09.doc

From: Pickett, Justin

Sent: Thursday, June 6, 2019 4:24 PM

To: bjohnso2@umd.edu; Eric Stewart; Gertz, Marc

Subject: Files and Concerns

Brian, Eric and Marc,

I have spent the day going back through all of my records for our 2011 Criminology article. I located files and emails that, without additional data, I have difficulty attributing to any benign explanation.

Here is the background: the data for our 2011 paper were collected in early 2008, and the paper was written in 2009. In fall 2009, after the analysis was finished and the paper was written, we sent it out for feedback from colleagues. ████████████████████████ ██████████████████████████ I was a graduate student at the time. I have attached the limited data Eric sent me (justin_voting_data.xls), as well as the data I sent back (Voting – Data.xls). Three things concern me.

████████████████████████████████ ████████████████████████████████. Eric sent me this data in December 2009, the same month we sent the manuscript to Criminology. ████████████████████ ██████████████

███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████

███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████ ███████████████████████████████████

I initially thought this must be the wrong data, even though Eric sent it to me for our paper. ███████████████████████████████████ ███████████████████████████████████

██ Thus, the variable's descriptives match the published statistics, even though the data are nothing like what we claim, ██

This led me to dig back through all of the emails I have from Eric in my old FSU account to see if I missed something. I found two emails that worry me. The first is his response when I apparently asked him (on September 11, 2010) ████████████████████████ He said:

██
██
██

At the time, as a graduate student, I found this explanation convincing. Now, I do not, for several reasons.

██
██
██
██

In another email, Eric sent me the 2009 ASC presentation for our paper (attached). I opened it and looked at the data and findings. ██
██
██
██
██

I remain hopeful that one of you will share a copy of the data with me that clears up these issues. They appear quite serious to me. Without additional data and a convincing explanation, I will have to retract my name from the article.

Best,
Justin

Justin T. Pickett
Associate Professor
School of Criminal Justice
University at Albany, SUNY

135 Western Avenue
Albany, NY 12222
Phone: 518-442-5224
Email: Jpickett@albany.edu

Peer Panel Document

Re: Eric Stewart

#3

Dear Criminology Editors:

My co-authors and I have been asked to provide responses to questions raised by critics of findings reported in Stewart, Mears, Warren, Baumer, and Arnio (2018, *Criminology*) and a published corrigendum to that paper. As the first author and the primary analyst for the study, I (Eric Stewart) am taking the lead in preparing this response. However, all co-authors have had an opportunity to review and approve the response. Below are responses to the questions that have been raised, along with additional information that we hope will clarify the concerns that have been raised. This memo highlights that the published works are not based on fabricated or falsified data or analysis. It also highlights that there were several coding errors, some mistakes in transcribing tabular results from statistical output to tables submitted for publication, and a need for more detail about the data and methods applied. The following eight points summarize these matters.

(1) Questions have been raised about the source data used for the paper. For your records, I have included a copy of an email that ███ ██ ███████████████████████████████████ For the Stewart et al. (2018, *Criminology*) study, we relied on public opinion data that were received from the Research Network in 2013. It was incorrectly assumed and reported that the data were collected in 2013, but Dr. Gertz has since clarified that while the data used for the study were distributed to me in 2013, these data were collected between January 2008 and December 2010. We regret that this was not explained in the published paper or corrigendum.

(2) In the original manuscript, the standard deviations for some of the binary measures were incorrect. This is because the paper mistakenly reported the standard deviations of those variables in their original ordinal scales rather than the recoded binary versions of these measures used in the study (the different versions of the variables had very similar names, and unfortunately the wrong versions were included in the descriptive statistics reported in the original paper). The corrigendum does not report standard deviations for the binary measures because, unlike standard deviations for continuous measures, they are not readily interpretable and do not provide information unique to the sample.

(3) In the original manuscript, all coefficients and standard errors were ███████████████ ██ ██ did not impact the substantive conclusions reported in the paper but should have been explained. The results reported in the corrigendum reflect no rounding for the coefficients and standard errors.

(4) The anonymous critique of the corrigendum highlights that it is not mathematically possible that removing the counties and 140 respondents sampled within them could result in the observed differences in mean age reported in the original paper and the corrigendum. That is correct. This ambiguity noted by this critic resulted from a transcription error—age was not accurately recorded in table 1 of the original paper. This error was corrected in the corrigendum, which correctly reports the mean age of 47.

(5) In preparing this response, I discovered that the gender variable (male) was inaccurately reported in the corrigendum because of a recoding error. I apologize. The original variable was coded as 2 = males, 1 = females. I recoded the variable to a 1/0 measure, with 1 = males and 0 = females, and separately had a version that was coded as 0 = males and 1 = females. The latter version was inadvertently reported in the paper, but referenced as percentage male. This mistake has subsequently been corrected, as shown in the output. Males represent 43 percent of the sample while females represent 57 percent of it. These percentages are consistent with other studies published in *Criminology* that have used other survey data collected by the Research Network, including articles by Chiricos et al. (2004) and Pickett (2012). I also identified that in table 2, model 4, the coefficient for homicide rate is mistakenly listed as -.014 in the corrigendum, when it should be -.044.

(6) In the original version and corrigendum to the manuscript, the source data for the income measure was a 6-category ordinal measure, which ranged from "1 = lowest income category" to "6 = highest income category." The original manuscript and corrigendum should have included more detail to explain that each of the six categories reflected income ranges (e.g., 1 = less than $15,000 to 6 = $100,000 or more). To provide readers with a more intuitive sense of income differences, we used the (midpoint) income value of these categories to represent subject income; the reported means and standard deviation were based on these values. Despite the corrections made to the original, the mean income (and the corresponding midpoint) remained unchanged.

(7) As noted in the corrigendum, the main substantive revision to the original article consisted of removing respondents who resided in non-southern counties from analyses reported in tables 1-3 and figures 1 and 2. Specifically, when reviewing the analysis after questions were raised about the results reported in the original article, I discovered that some counties had been incorrectly recorded as "southern." Removing these counties yielded a sample that was smaller both in terms of the number of counties and the number of respondents. That was the main substantive change to the data, which was discussed in the corrigendum.

Given the ensuing confusion about the comparison of results reported in the original paper and the corrigendum, it is clear that it would have been best to acknowledge two other differences between these versions. These are discussed below.

First, in the original version of the manuscript, the lynching counts were reported to be based on modern county and county cluster boundaries. However, after further investigation, I discovered that some of the boundaries applied in the data did not represent the preferred modern county boundaries. Specifically, in the original paper approximately 20 percent of the lynching counts were situated within historical county boundaries used in the late 1800s and early 1900s, rather than the modern boundaries. To correct this discrepancy, I reconstructed the historical county-level lynching data to uniformly correspond to contemporary boundaries and re-estimated all analyses reported in the paper. The corrigendum to the *Criminology* article reflects this correction. I regret not describing this modification in the corrigendum; I had decided to not do so because it had no bearing on the substantive conclusions drawn in the study, though it did contribute to some changes to the descriptive statistics, coefficients, and standard errors reported in the corrigendum.

Second, the results presented in the corrigendum did not address, through imputation, missing data for some measures; by contrast, imputation was used in the analyses reported in the original version of the paper. That difference in method was an oversight and is the key reason why some of the values in the original and the corrigendum were not directly comparable. The statistical and substantive results in the two were very similar. Accordingly, not using imputation with the corrigendum did not affect the main conclusions that were reported in the published paper or the corrigendum.

Even so, to create a more direct comparison, I re-ran the analyses from the corrigendum using imputed data. There was no significant substantive or statistical difference in the two sets of results. I then revisited this analysis after introducing two minor changes—I used the percent voting Republican in the 2008 presidential election rather than the 2012 election, given that the public opinion poll occurred in 2010 (as explained in point 1 above) and corrected the coding for the gender variable (as explained in point 5 above). After re-running the corrigendum analyses using imputation with these two changes included, the main conclusions were, again, substantively and statistically similar to those reported in the original and corrigendum. For supporting evidence, please see the accompanying output, which relied on analysis of imputed data that incorporated the corrected gender variable and the updated indicator of percent Republican.

(8) Questions have been raised about the means and ranges reported in the corrigendum for concentrated disadvantage, population structure, and black population growth. Specifically, the maximum values reported in the corrigendum for these variables are greater than the maximums reported in the original manuscript. For concentrated disadvantage, this occurred because we added a larger constant (13 compared to 12) to rescale the variable as a positive integer. For population structure and black growth, this occurred because of the reconfigured county boundaries, which yielded areas with higher upper-range values.

We hope that these responses clarify questions that have been raised about the data and results published in the original manuscript and the corrigendum. As we have outlined above, the discrepancies that have been noted are a function of some unfortunate coding and transcription errors, along with insufficient information provided in the corrigendum about the data and corrections implemented. These had no bearing on the substantive conclusions drawn. Even so, we regret the mistakes, and in retrospect it is clear that a more detailed explanation of corrections made between the original version of the paper and the corrigendum could have been instrumental in heading off additional concerns.

I would like to close by stating that there was and is no fabrication or falsification of data or results in this paper or any other papers I have authored. I feel compelled to make such a statement because, several months ago, an anonymous person contacted you and the editors of two other journals to highlight questions about data and results of five separate papers, with the strong implication being that I had engaged in fraud. Subsequently, a former co-author of one of the papers (2011, *Criminology*) reinforced this view after analyzing an incomplete data file and posting an on-line critique that, again, implied fraudulent research practices on my part. This person then shared that document with an online blog dedicated to chronicling instances of

research fraud (Retraction Watch) and with the editors of the three journals, used it as the basis of a complaint filed with Florida State University, submitted a public records request for my emails, and hired an attorney to assist him with further open records requests. A reporter from the *Chronicle of Higher Education* subsequently sought to contact me about assumed problems with the papers, based on discussions with the former co-author, and also filed a public records request for emails from the Dean of my College, the FSU Office of Research Integrity, and me. There is, in addition, the latest anonymous email sent to you raising questions about the corrigendum to the 2018 paper published in *Criminology*. This process has created an atmosphere in which discrepancies in research have been incorrectly assumed to reflect fraud rather than unintended and regrettable mistakes and oversights. Although some good has come from this process—for example, it led to the identification of mistakes with the original manuscript—it is unfortunate that non-reviewed and anonymous critiques contribute to such an atmosphere.

On behalf of my colleagues, I would like to thank you for your understanding as we worked through the concerns raised about the original 2018 publication and the corrigendum. If you would like additional information, please do not hesitate to let us know.

Sincerely,

Eric A. Stewart

Enc: (1) Email from Dr. Marc Gertz
       (2) Corrigendum output
       (3) Revised corrigendum output using imputation

Peer Panel Document

Re: Eric Stewart

#4

For you records

^^^^^^^^^^^^^^^^^^^^^^^^^
Gary K. Ostrander Ph.D.
Vice-President for Research
Florida State University
3012 Westcott North
Tallahassee, FL 32306-1330
850-644-3347 office
850-645-0108   FAX
gary@fsu.edu
^^^^^^^^^^^^^^^^^^^^^^^^^

From: tgblomberg@aol.com <tgblomberg@aol.com>
Sent: Thursday, October 24, 2019 7:18 PM
To: Gary Ostrander <Gary@fsu.edu>
Subject: Fwd: Withdrawal of 2018 Article

Gary,

fyi.

Tom

Thomas G. Blomberg
Dean and Sheldon L. Messinger Professor of Criminology
College of Criminology and Criminal Justice
Florida State University
Tallahassee, Florida 32306-1127
(850) 644-7380

-----Original Message-----
From: McDowall, David <dmcdowall@albany.edu>
To: Eric Stewart <estewart2@fsu.edu>; Lauritsen, Janet L. <janet_lauritsen@umsl.edu>;
jody.miller@rutgers.edu <jody.miller@rutgers.edu>
Cc: Daniel Mears <dmears@fsu.edu>; Patricia Warren-Hightower <pwarren@fsu.edu>; ERIC P
BAUMER <epb5167@psu.edu>; Arnio, Ashley <aarnio@txstate.edu>; Thomas Blomberg
<tblomberg@fsu.edu>
Sent: Thu, Oct 24, 2019 7:05 pm
Subject: RE: Withdrawal of 2018 Article

Hi Eric,

My co-editors and I have discussed your email, and we will move forward with the
retraction. We will get back to you with the details of how to proceed as we consult with the
publisher and the ASC Publications Committee. I know that it must have been a hard and
painful decision to make, but it is also a courageous decision and reflects well on your

commitment to scholarship. For what it is worth, you have my respect.

David

David McDowall
School of Criminal Justice
University at Albany—SUNY
135 Western Avenue
Albany, NY   12222

Voice: 518-442-5225

From: Eric Stewart <estewart2@fsu.edu>
Sent: Thursday, October 24, 2019 2:05 AM
To: McDowall, David <dmcdowall@albany.edu>; Lauritsen, Janet L.
<janet_lauritsen@umsl.edu>; jody.miller@rutgers.edu
Cc: Daniel Mears <dmears@fsu.edu>; Patricia Warren-Hightower <pwarren@fsu.edu>;
ERIC P BAUMER <epb5167@psu.edu>; Arnio, Ashley <aarnio@txstate.edu>; Thomas
Blomberg <tblomberg@fsu.edu>
Subject: Withdrawal of 2018 Article

Dear Professors McDowall, Lauritsen, and Miller:

My co-authors and I request withdrawal of Stewart et al. (2018, Lynchings, Racial
Threat, and Whites' Punitive Views towards Blacks, *Criminology*). Contrary to
assertions in social media and other online outlets, there was no falsification,
fabrication, or other fraud associated with this paper or others that relied on the public
opinion and contextual data. However, questions have been raised about the data
and results reported in the paper. I have taken the lead in responding to these
questions because I undertook the analyses, but this request for withdrawal comes
from all of the co-authors. Addressing the questions has yielded findings that
affirmed the main substantive results. Nonetheless, in the course of responding to the
questions, it became clear that the errors reflected in the paper, including coding
mistakes and transcription errors, exceeded what the co-authors and I view as
acceptable for a published paper. For these reasons, on behalf of the co-authors I
ask that the paper be withdrawn.

Sincerely,

Eric Stewart, Ph.D.
Florida State University

Peer Panel Document
Re: Eric Stewart

#5

Gary,

FYI,

Tom

Sent from my iPhone

Begin forwarded message:

> **From:** Eric Stewart <estewart2@fsu.edu>
> **Date:** October 24, 2019 at 2:01:49 AM EDT
> **To:** Susan Sterett <ssterett@umbc.edu>, Law & Society Review <lsr@umbc.edu>
> **Cc:** Daniel Mears <dmears@fsu.edu>, Patricia Warren-Hightower <pwarren@fsu.edu>, MiltonetteCraig <mocraig@ilstu.edu>, "Arnio, Ashley" <aarnio@txstate.edu>, Thomas Blomberg <tblomberg@fsu.edu>
> **Subject: Withdrawal of 2019 Article**
>
>
> Dear Professors Jeannine Bell, Susan M. Sterett, and Margot Young:
>
> My co-authors and I request withdrawal of Mears et al. (2019, "A Legacy of Lynchings: Perceived Black Criminal Threat among Whites", *Law and Society Review*). Contrary to assertions in social media and other online outlets, there was no falsification, fabrication, or other fraud associated with this paper or others that relied on the public opinion and contextual data. However, questions have been raised about the data and results reported in the paper. I have taken the lead in responding to these questions because I undertook the analyses, but this request for withdrawal comes from all of the co-authors. Addressing the questions has yielded findings that affirmed the main substantive results. Nonetheless, in the course of responding to the questions, it became clear that the errors reflected in the paper, including coding mistakes and transcription errors, exceeded what the co-authors and I view as acceptable for a published paper. For these reasons, on behalf of the co-authors I ask that the paper be withdrawn.
>
>
> Sincerely,
>
> Eric Stewart, Ph.D.
> Florida State University

Peer Panel Document
Re: Eric Stewart

#6

Dear all,

This is just to notify you that the retractions are now posted on OUP's journal site.

This is what the note says:

"The authors of "…..TITLE…." have retracted this article because of errors uncovered in the paper subsequent to its publication that exceed what the authors view as acceptable for a published paper."

Anna

From: Eric Stewart [mailto:estewart2@fsu.edu]
Sent: Friday, October 25, 2019 5:01 PM
To: Wright, Earl (wrighte4) <wrighte4@UCMAIL.UC.EDU>; Linders, Annulla (lindera) <lindera@ucmail.uc.edu>
Cc: Patricia Warren-Hightower <pwarren@fsu.edu>; Brian D Johnson <bjohnso2@umd.edu>; Marc Gertz <mgertz@fsu.edu>; Cresean Hughes <cresean@udel.edu>; ERIC P BAUMER <epb5167@psu.edu>; Martinez, Ramiro <R.Martinez@northeastern.edu>; Rosario, Jordyn <jlr10d@my.fsu.edu>; Thomas Blomberg <tblomberg@fsu.edu>
Subject: Withdrawal of 2019 & 2015 Articles

Dear Professors Linders and Wright:

My co-authors and I request withdrawal of Stewart et al. (2019, The Social Context of Criminal Threat, Victim Race, and Punitive Black and Latino Sentiment, *Social Problems*) and (2015, The Social Context of Latino Threat and Punitive-Latino Sentiment, *Social Problems*). Contrary to assertions in social media and other online outlets, there was no falsification, fabrication, or other fraud associated with these paper or others that relied on the public opinion and contextual data. However, questions have been raised about the data and results reported in these papers. I have taken the lead in responding to these questions because I undertook the analyses, but this request for withdrawal comes from all of the co-authors. In the course of responding to the questions, it became clear that the errors reflected in the papers, including coding mistakes and transcription errors, exceeded what the co-authors and I view as acceptable for a published paper. For these reasons, on behalf of the co-authors I ask that the papers be withdrawn.

Sincerely,

Eric Stewart, Ph.D.
Florida State University

Peer Panel Document

Re: Eric Stewart

#7

On January 21, 2021, the Editors and Publisher of *Justice Quarterly* retracted Dr. Stewart's sole-authored 2003 paper.

   Stewart, Eric A. 2003. School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly* 20:575-604.

Their published retraction commentary read in-part "*In response to concerns raised about this article, the Editors and the Publisher provided the author with the opportunity to respond to the questions raised about this article, and commissioned a review by three independent referees of the original article and the author's response to the concerns.*

Owing to the significant length of time that has passed since this article was published, the author did not retain the data or analyses used in this paper. The author and referees accessed the publicly available dataset, the first follow-up to the National Educational Longitudinal Study, from the National Center for Education Statistics.

The independent review concluded that some of the concerns raised about the article could possibly be attributed to analytic and coding decisions, but due to possible errors in select analyses, as well as the inability of the referees to replicate the article's exact sample and models, there was doubt cast over the reliability of the results and conclusions in this study. For this reason, we have made the decision to retract the article. The author has been informed of this decision.

Based upon the information from the author and the independent review, there was insufficient evidence to support a conclusion of research misconduct in the form of data fabrication and falsification."

Peer Panel Document

Re: Eric Stewart

#8

Reanalysis Report:
Johnson, Stewart, Pickett, and Gertz (2011)

Justin T. Pickett
School of Criminal Justice
University at Albany, SUNY

ABSTRACT

My coauthors and I were informed about data irregularities in Johnson, Stewart, Pickett, and Gertz (2011), and in my coauthors' other articles. Subsequently, I examined my limited files and found evidence that we: 1) included hundreds of duplicates, 2) underreported the number of counties, and 3) somehow added another 316 respondents right before publication (and over a year after the survey was conducted) without changing nearly any of the reported statistics (means, standard deviations, regression coefficients). The survey company confirmed that it sent us only ▇▇▇▇▇▇▇▇▇, not the 1,184 reported in the article. I obtained and reanalyzed those data. This report presents the findings from my reanalysis, which suggest that the sample was not just duplicated. The data were also altered—intentionally or unintentionally—in other ways, and those alterations produced the article's main findings. Additionally, we misreported data characteristics as well as aspects of our analysis and findings, and we failed to report the use of imputation for missing data. The following eight findings emerged from my reanalysis:

1) The article reports 1,184 respondents, but actually there are ▇▇▇
2) The article reports 91 counties, but actually there are ▇▇
3) The article describes respondents that differ substantially from those in the data.
4) The article reports two significant interaction effects, but actually ▇▇▇▇▇▇.
5) The article reports the effect of Hispanic growth is significant and positive, but actually it ▇▇▇▇▇▇▇▇▇▇▇▇▇▇▇▇.
6) The article reports many other findings that do not exist in the data.
7) The standard errors are stable in our published article, but not in the actual data or in articles published by other authors using similar modeling techniques with large samples.
8) Although never mentioned in the article, 208 of the ▇▇ respondents in the data (or ▇▇▇ have imputed values.

---

*Direct correspondence to Justin T. Pickett, School of Criminal Justice, University at Albany, SUNY, 135 Western Avenue, Albany, NY 12222 e-mail (jpickett@albany.edu).

BACKGROUND

On May 5, 2019, my coauthors and I received an email that identified data irregularities in our study, Johnson, Stewart, Pickett, and Gertz (2011), and in four other articles authored or coauthored by Dr. Stewart. I asked Dr. Stewart to send me the full data for Johnson et al. (2011), but encountered difficulties getting them. Consequently, I examined the limited files I already had, and ████████████████████████████████████, as well as several other problems, like inexplicable changes in sample size from manuscript draft ███████ to published article ($N = 1,184$) that did not affect means, standard deviations, or regression coefficients. I sent an email to my coauthors on June 6 that listed these issues and provided my files (see Appendix A). One of my coauthors, Dr. Gertz, then contacted the former director of the Research Network, who confirmed that the survey he ran for us included only ██ respondents (see Appendix B). At that point, Dr. Stewart sent me a copy of the data for our article (████████

In the article, we claim to have 1,184 respondents nested in 91 counties. In the actual data, there are only ██ respondents, and they are nested in ██ counties. Dr. Stewart acknowledged that both the sample size and county number reported in the article were wrong. He said the explanations for the differences were that: ████████████████████████████████ ███. The published article never mentions ██████████████████████████████. It is also unclear how the sample of ██ grew to 1,184 in our article, and then to 1,379 in Dr. Stewart's later *Social Problems* article (Stewart et al., 2015), which uses the same data (with the same 54.8% response rate, and same $62,700 mean family income). ████████████████████████ ██████████████████████████

Dr. Stewart now says ███████████████████████████████████████

███ are similar

to those in the published article. However, I am uncomfortable with the new results for four

reasons. First, I have not seen them. Dr. Stewart not sent me the data (or even the Stata output)

for the second sample. Second, the published article reports 1,184 respondents, not ███ Third,

our published article lists only one survey company—the Research Network—and one survey.

Fourth, ███ .

Before I discovered the duplicates in our data, ███████████████████████

███████████████████████████████████████ The email from the Research

Network (see Appendix B) says it conducted one ████████████████████████████

███

███████████████████████████████████████Therefore, the remainder of

this report focuses on the findings from my reanalysis of the data I can account for ██████████

which are from the sample that our article describes (albeit inaccurately) and that Dr. Stewart

██████████████████████████ . I did ask Dr. Stewart for the full sample (with duplicates

included) of 1,184 respondents that he initially said he used in his original analysis for our

article. Unfortunately, he doesn't have it, because he saved over it after dropping the duplicates.

REANALYSIS FINDINGS

The evidence suggests the data were altered in ways besides duplication. The descriptive

statistics in the published article should match ███████████████████████████████████

███

████████████ . However, the descriptive statistics in the published article differ substantially

from the actual data. The outcome variable in our analysis is public support for the use of defendants' ethnicity in sentencing decisions. The distribution of the outcome variable by respondents' race is shown in Figure 1 of the published article (Johnson et al., 2011: 419). Here is how it compares to the actual data:

Figure 1.   Published Article Vs. Actual Data
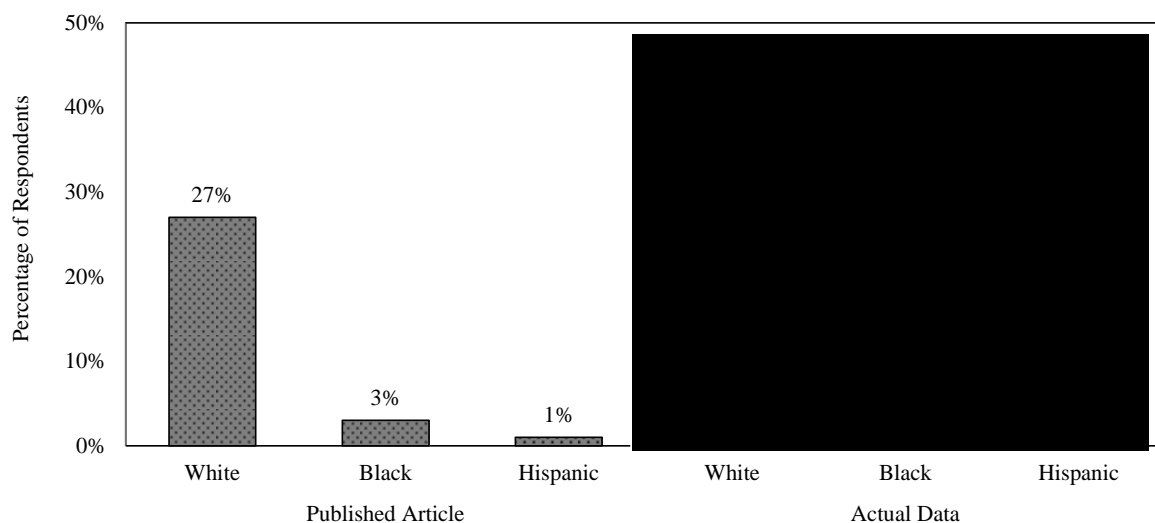
cannot explain these discrepancies. For example, even adding ▮ Black respondents who ALL oppose ethnicity-based sentencing would reduce the percent of Blacks supporting it from 38% to 13%, not to the article's 3%. And this is not the only noteworthy distributional difference. The table below compares all of the descriptive statistics in the published article to those in the actual data. The means for ▮ of the ▮ variables (or ▮ differ significantly between the article and data. For example, the published article claims that 43% of respondents are political conservatives. In the actual data, ▮ are political conservatives.

Again, even ███████████████████████████████████████, would only drop the

percentage of conservatives in the sample to 59%, not to the 43% reported in the article.

Table 1.   Descriptive Statistics: Published Article vs. Actual Data

Variables	Published Article		Actual Data
	Mean	SD	
Use of ethnicity in punishment	.31	.46	
Hispanic criminal threat	4.93	1.66	
Hispanic economic threat	1.72	1.13	
Hispanic political threat	4.38	1.41	
Percent Hispanic	.12	.11	
Hispanic growth	.26	1.53	
Homicide rate (per 100,000)	3.96	4.37	
Concentrated disadvantage	1.09	1.53	
Percent Republican	53.04	13.02	
Percent Black	.10	.14	
Population structure	5.39	.70	
White	.86	.41	
Black	.10	.33	
Hispanic	.04	.22	
Age	47.12	19.72	
Male	.47	.50	
Married	.59	.31	
Education level (college graduate)	.42	.31	
Family income	$62,700	$14,210	
Employed	.46	.50	
Political conservative	.43	.31	
Own home	.78	.33	
Southwest	.17	.41	
Northeast	.15	.35	
Midwest	.24	.43	
West	.17	.38	
South	.44	.39	
General punitive attitudes	6.84	2.16	

Similarly, the mean for Hispanic criminal threat is almost ███████████████████

███████ and the mean for general punitive attitudes is over ████████████████████████

███), in the published article than in the actual data. ██████████████████████████

██

███████████████████████████████; both still a point lower than in the article.

██ would leave the regression coefficients unscathed—they would be identical if all respondents ████████████ However, the regression results in the published article differ substantially from those in the actual data. Most notably, the main findings in the article—the interaction effect of perceived Hispanic threat (criminal and economic) and Hispanic growth—███████████████████████████ Those findings are reported in Table 3 of Johnson et al. (2011: 423). In the actual data, ████████ ██.

This is shown in the table below.

Table 2. Interaction Effects: Published Article vs. Actual Data

	Published Article			Actual Data		
Variables	*b*	SE	Exp(b)	████		
Variables	—	—	—	████		
Perceived Hispanic threat	—	—	—	████		
Criminal threat	.183*	.079	1.201	████		
Economic threat	.272*	.111	1.312	████		
Political threat	.008	.116	1.009	████		
Aggregate Hispanic threat	—	—	—	████		
Percent Hispanic	−.089	.766	.815	████		
Hispanic growth	.334**	.127	1.396	████		
Interactions	—	—	—	████		
Criminal * His. Growth	.126*	.051	1.134	████		
Economic * His. Growth	.175*	.073	1.191	████		
Political * His. Growth	−.101	.087	.904	████		
Intercept	−.848***	.119	—	████		

 *p < .05; **p < .01; ***p < .001 (two-tailed).

The main effect of Hispanic growth also fails to replicate in the actual data; indeed, the coefficient is in████████████████████. This is the case even when the interaction terms are removed from the model. In the published article, the main effect of Hispanic growth is shown Model 2 of Table 2, and is positive and highly statistically significant (*b* = .288, *p* < .01). In the actual data, ██████████████████████████. The table below compares the

estimates in Johnson et al. (2011: 420) to those from the actual data. The differences are striking, extending to many other variables besides Hispanic growth. For example, ██████████████

██

██

Table 3.   Model 2 in Published Article vs. Actual Data

Variables	Published Article			Actual Data		
	b	SE	Exp(b)	███		
Percent Hispanic	−.104	.662	.901			
Hispanic growth	.288**	.102	1.334	█		
Homicide rate (per 100,000)	−.013	.033	.987			
Concentrated disadvantage	.051	.086	1.052			
Percent Republican	.001	.005	1.000			
Percent Black	−.038	.173	.963			
Population structure	−.051	.174	.950			
Black	−.628*	.248	.534			
Hispanic	−.982**	.359	.375			
Age	.016**	.005	.016			
Male	.164**	.056	1.178			
Married	.208	.201	1.231			
Education level	−.099	.198	.906			
Family income	−.026	.061	.974			
Employed	−.191*	.082	.826			
Political conservative	.367*	.146	1.443			
Own home	−.137	.246	.872			
Southwest	−.111	.274	.895			
Northeast	−.183	.281	.832			
Midwest	.054	.230	1.055			
West	.082	.264	1.085			
General punitive attitudes	.191**	.062	1.210			
Intercept	−.858***	.105	—			

*p < .05; **p < .01; ***p < .001 (two-tailed).

There are other issues with the data that are concerning. For example, one of the irregularities raised in the email we received was the high degree of stability in standard errors across the three models in Table 2 of our published article (Johnson et al., 2011: 420-421). In the article, 21 regressors are included in all three models and 19 of them (or 90%) have standard errors that are perfectly stable. In the actual data, ████████████████████████████████

As the table below shows, this is the case regardless of whether the models are estimated using logistic regression (with clustered standard errors) or multilevel modeling. In the table, there are boxes around the stable standard errors.

Table 4.   Standard-Error Stability: Published Article vs. Actual Data

Variables	Published Article			Actual Data: Logistic	Actual Data: Multilevel
	Model 1: SE	Model 2: SE	Model 3: SE		
Criminal threat	—	—	.084		
Economic threat	—	—	.111		
Political threat	—	—	.107		
Percent Hispanic	—	.662	.669		
Hispanic growth	—	.102	.101		
Homicide rate	.033	.033	.033		
Concentrated dis.	.086	.086	.086		
Percent Repub.	.005	.005	.005		
Percent Black	.173	.173	.173		
Population structure	.173	.174	.185		
Black	.248	.248	.248		
Hispanic	.359	.359	.359		
Age	.005	.005	.005		
Male	.056	.056	.056		
Married	.201	.201	.201		
Education level	.198	.198	.198		
Family income	.061	.061	.061		
Employed	.082	.082	.082		
Political conservative	.146	.146	.146		
Own home	.246	.246	.246		
Southwest	.274	.274	.274		
Northeast	.281	.281	.281		
Midwest	.230	.230	.230		
West	.264	.264	.264		
General punitive	.062	.062	.062		
Intercept	.102	.105	.119		

NOTES: Stable standard errors are boxed in.

The observed differences in standard-error stability between our published article and the actual data are so startling that I searched for other articles to compare. I found several published by other prominent scholars in top journals that, in design and analysis, are comparable to ours. Specifically, they all have large samples, include a series of multilevel regression models

8

examining one outcome variable (stepwise, building from a baseline model), and report standard errors to the third decimal place:

- ➢ Hagan, Shedd, and Payne (2005), *American Sociological Review*
- ➢ Kirk (2008), *Demography*
- ➢ Kirk and Matsuda (2011), *Criminology*
- ➢ Sampson, Morenoff, and Raudenbush (2005), *American Journal of Public Health*
- ➢ Slocum, Taylor, Brick, and Esbensen (2010), *Criminology*
- ➢ Xie, Lauritsen, and Heimer (2012), *Criminology*

I have included the relevant regression tables from all of these articles in Appendix C of this report. None of the articles exhibit the degree of standard-error stability we report in Johnson et al. (2011). Instead, they all are consistent with the actual data we have, and show that standard errors normally vary across stepwise multilevel models, the main exception being standard errors with two leading zeros (e.g., SE = .005).

Another issue with our data is item missing values. In the file Dr. Stewart sent me, all ▮▮▮ respondents ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮. In most surveys, a substantial number of respondents have missing values on some of the variables (e.g., income). Closer inspection of the data reveals that ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

CONCLUSION

There is only one possible conclusion from reanalyzing the data I have: the sample was not just ██████ in the analysis for the published article; ████████████████████ ██. It may be that appending the data I have and the data Dr. Stewart has for the second sample of ██ respondents would change this conclusion. Unfortunately, ████████████████ ████████████████████████████. Regardless, our published article did not report a second survey, or a sample of ██ it reported one survey of 1,184 respondents by the Research Network.

Three other things are incontrovertible. First, we omitted important information that must be reported to journal referees and readers, like the use of ██████████████████ Second, we misreported data characteristics, like the number of counties—████████████ ██████████Third, if Dr. Stewart ██████████████████████

██

██

██

██

██

██

██

---

[1] There seems to be little reason to use county clusters in our study. It is not desirable to group together counties, because it throws away geographic detail and creates "meaningless socio-political entities" (Hagen et al., 2013: 770). Typically, researchers only group together counties in historical studies that examine data over a large number of decades or across centuries, and even in those studies they only create county clusters for those specific counties that have boundaries that changed during the time period examined (e.g., King et al., 2009; Messner et al., 2005).

REFERENCES

Hagan, John, Carla Shedd, and Monique R. Payne. 2005. Race, ethnicity, and youth perceptions of criminal injustice. *American Sociological Review* 70: 381-407.

Hagen, Ryan, Kinga Makovi, and Peter Bearman. 2013. The influence of political dynamics on southern lynch mob formation and lethality. *Social Forces* 92: 757-787.

Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz. 2011. Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. *Criminology* 49: 401-441.

King, Ryan D., Steven F. Messner, and Robert D. Baller. 2009. Contemporary hate crimes, law enforcement, and the legacy of racial violence. *American Sociological Review* 74: 291-315.

Kirk, David S. 2008. The neighborhood context of racial and ethnic disparities in arrest. *Demography* 45: 55-77.

Kirk, David S., and Mauri Matsuda. 2011. Legal cynicism, collective efficacy, and the ecology of arrest. *Criminology* 49: 443-472.

Messner, Steven F., Robert D. Baller, and Matthew P. Zevenbergen. 2005. The legacy of lynching and southern homicide. *American Sociological Review* 70: 633-655.

Sampson, Robert J., Jeffrey D. Morenoff, and Stephen Raudenbush. 2005. Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health* 95: 224-232.

Slocum, Lee A., Terrance J. Taylor, Bradley T. Brick, and Finn-Aage Esbensen. 2010. Neighborhood structural characteristics, individual-level attitudes, and youths' crime reporting intentions. *Criminology* 48: 1063-1100.

Stewart, Eric A., Ramiro Martinez, Jr., Eric P. Baumer, and Marc Gertz. (2015) The social context of latino threat and punitive Latino sentiment. *Social Problems* 62: 69-92.

Xie, Min, Janet L. Lauritsen, and Karen Heimer. 2012. Intimate partner violence in U.S. metropolitan areas: The contextual influences of police and social services. *Criminology* 50: 961-992.

APPENDIX A: "FILES AND CONCERNS" EMAIL

Brian, Eric and Marc,

I have spent the day going back through all of my records for our 2011 Criminology article. I located files and emails that, without additional data, I have difficulty attributing to any benign explanation.

Here is the background: the data for our 2011 paper were collected in early 2008, and the paper was written in 2009. In fall 2009, after the analysis was finished and the paper was written, we sent it out for feedback from colleagues. ███████████████████████████ ██ ████████████████████████ I was a graduate student at the time. I have attached the limited data Eric sent me (justin_voting_data.xls), as well as the data I sent back (Voting – Data.xls). Three things concern me.

██ ████████████████████████████████. Eric sent me this data in December 2009, the same month we sent the manuscript to Criminology. ████████████████████████████ ████████████████

██ ██ ██ ██ ██ ██████████████████████████████

██ ██ ██ ██ ██ ██ ██████████████████████████

I initially thought this must be the wrong data, even though Eric sent it to me for our paper. ███ ██ ██ Thus, the variable's descriptives match the published statistics, even though the data are nothing like what we claim, ██.

This led me to dig back through all of the emails I have from Eric in my old FSU account to see if I missed something. I found two emails that worry me. The first is his response when I

apparently asked him (on September 11, 2010) ████████████████████████████████. He said:

██
██
██
████████████████████████████████

At the time, as a graduate student, I found this explanation convincing. Now, I do not, for several reasons. ██
██
██
██
██
██
████████████████████████████████████
████████████████████████████████████
██
██████████████████████████████████

In another email, Eric sent me the 2009 ASC presentation for our paper (attached). I opened it and looked at the data and findings. ████████████████████████████
██
██
██
████████████████████████████████████
██
████████████████████████████████████
██
████████████████████████████████
██████████████████████████████

I remain hopeful that one of you will share a copy of the data with me that clears up these issues. They appear quite serious to me. Without additional data and a convincing explanation, I will have to retract my name from the article.

Best,
Justin

From: Jake Bratton <jbratton@jakebratton.com>
Date: June 7, 2019 at 9:33:13 AM EDT
To: Marc Gertz <mgertz@fsu.edu>
Subject: Re: Files and Concerns
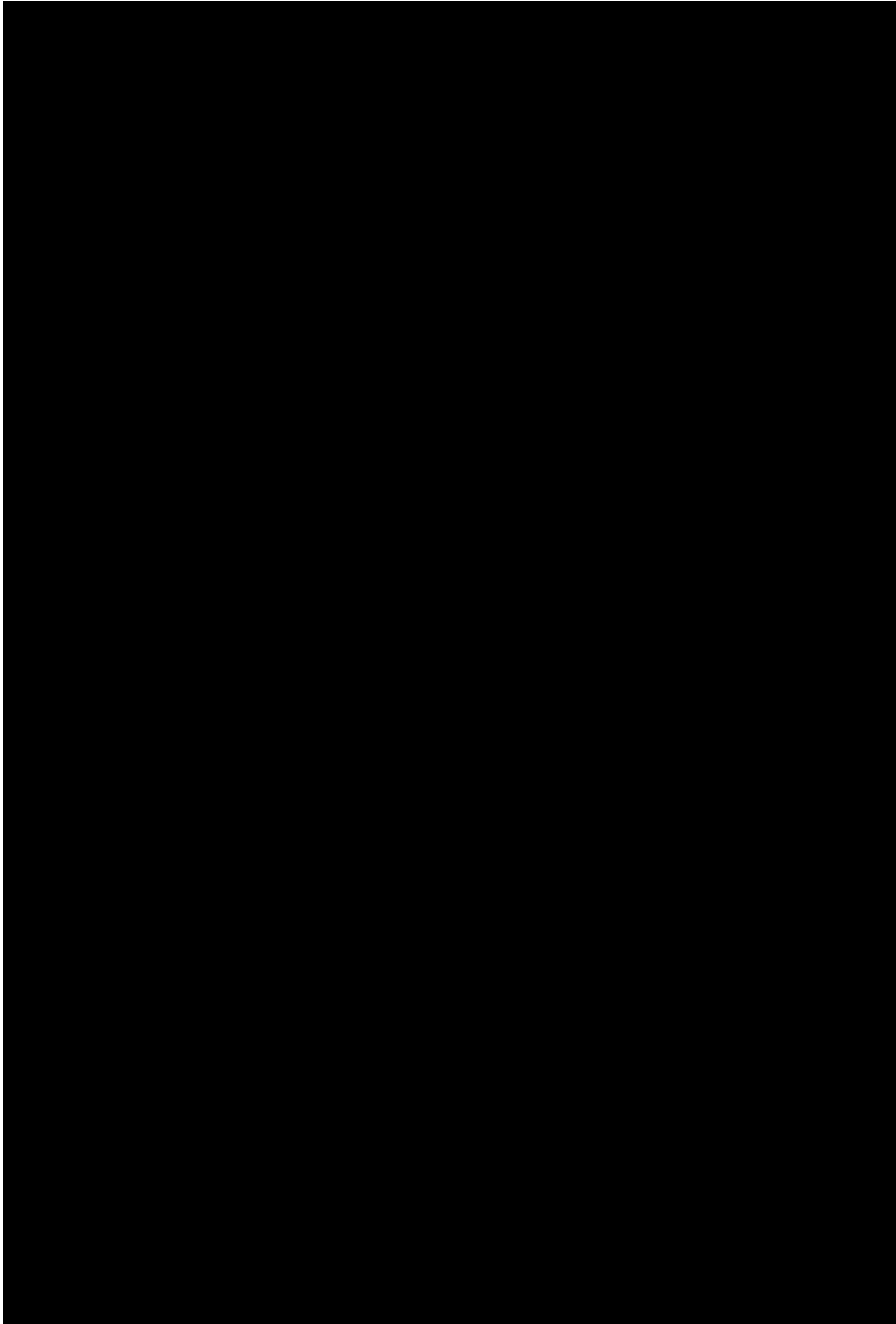
Just FYI, Dr. Stewart was sent exactly ████ records that January 2008 and a set of cross tabs of ████████ The demographics used in the methods section match the ██████ records he was sent for ██████████████████ exactly from those tabs I ran and distributed to four other professors around the same time. The county thing Eric said we messed up was that in May of 2008 I sent another spss file that only had the names/census info of the counties for the ██████████ and I think he had to would have looked up the county names by zip code in the file for the other ██████ hand to do his analysis. I don't have a record of providing the county level info for those other ██████ or so, or even that the file I sent was incomplete.
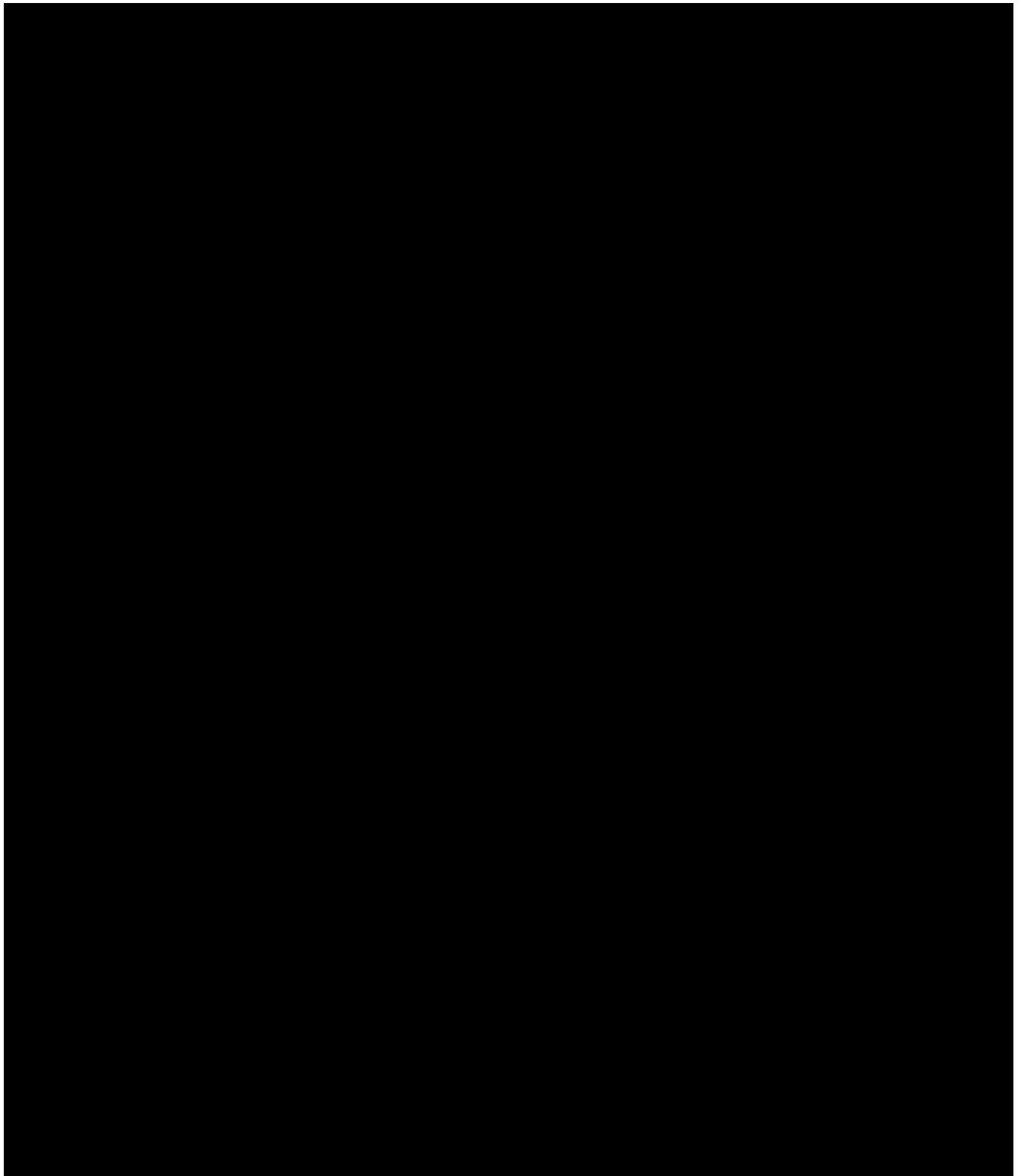
Basically the choices are inexperience in blending data resulting in loading the data twice and not noticing or more sample size was needed to get to p<.05.
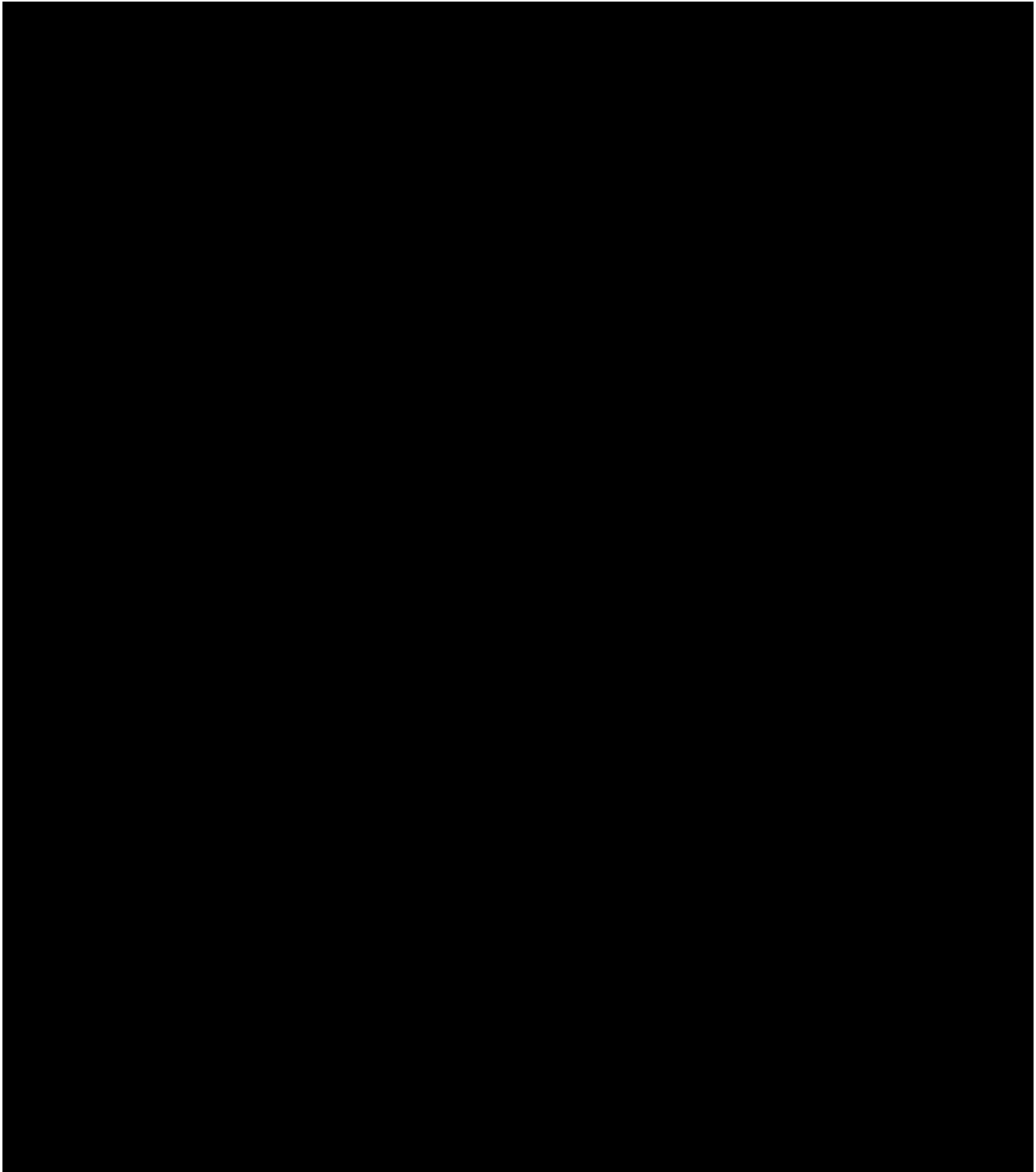
Table 3. Impact of School Racial Composition on White Youth Perception of Injustice
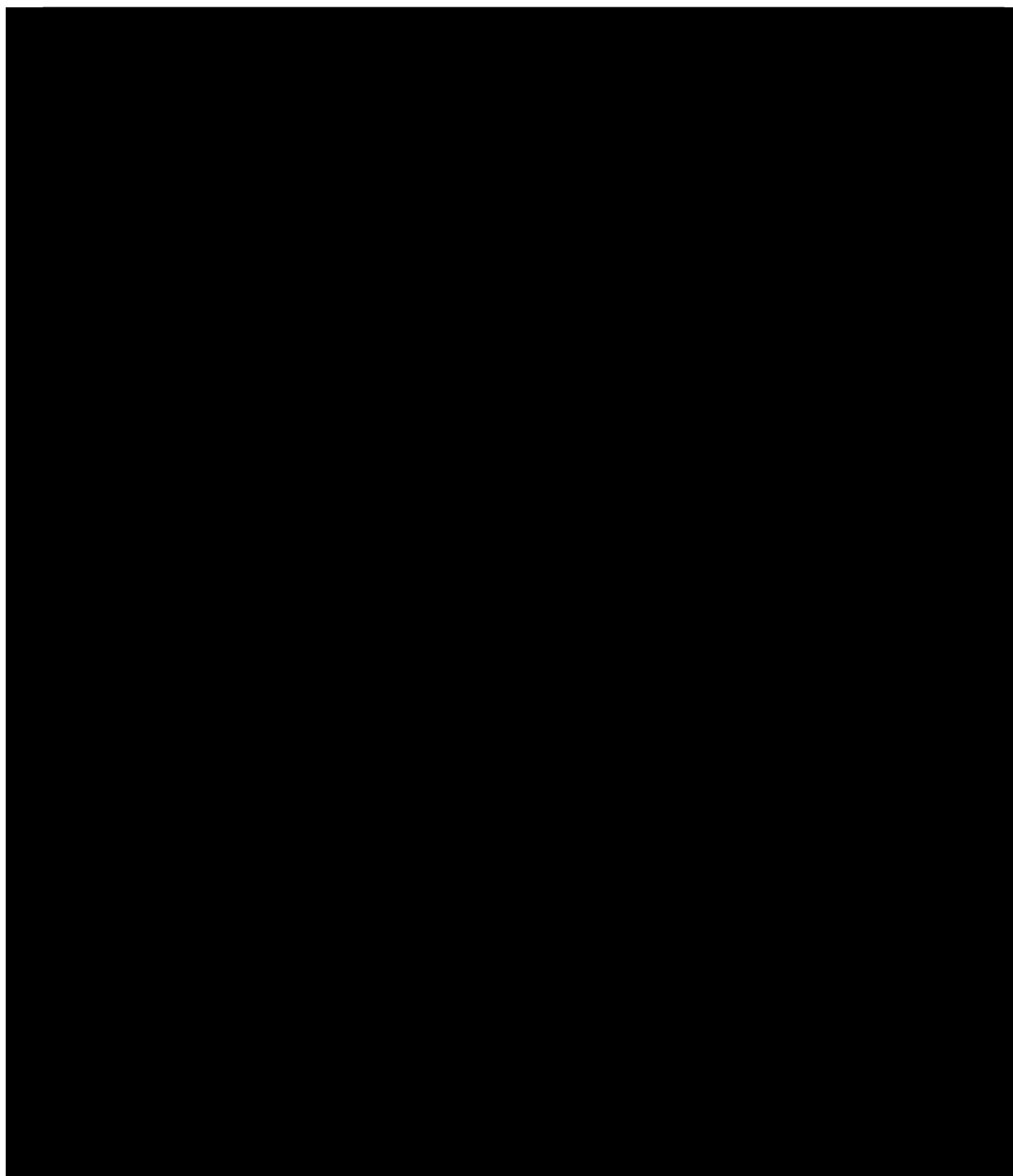
Table 4. Impact of School Racial Composition on Black Youth Perception of Injustice
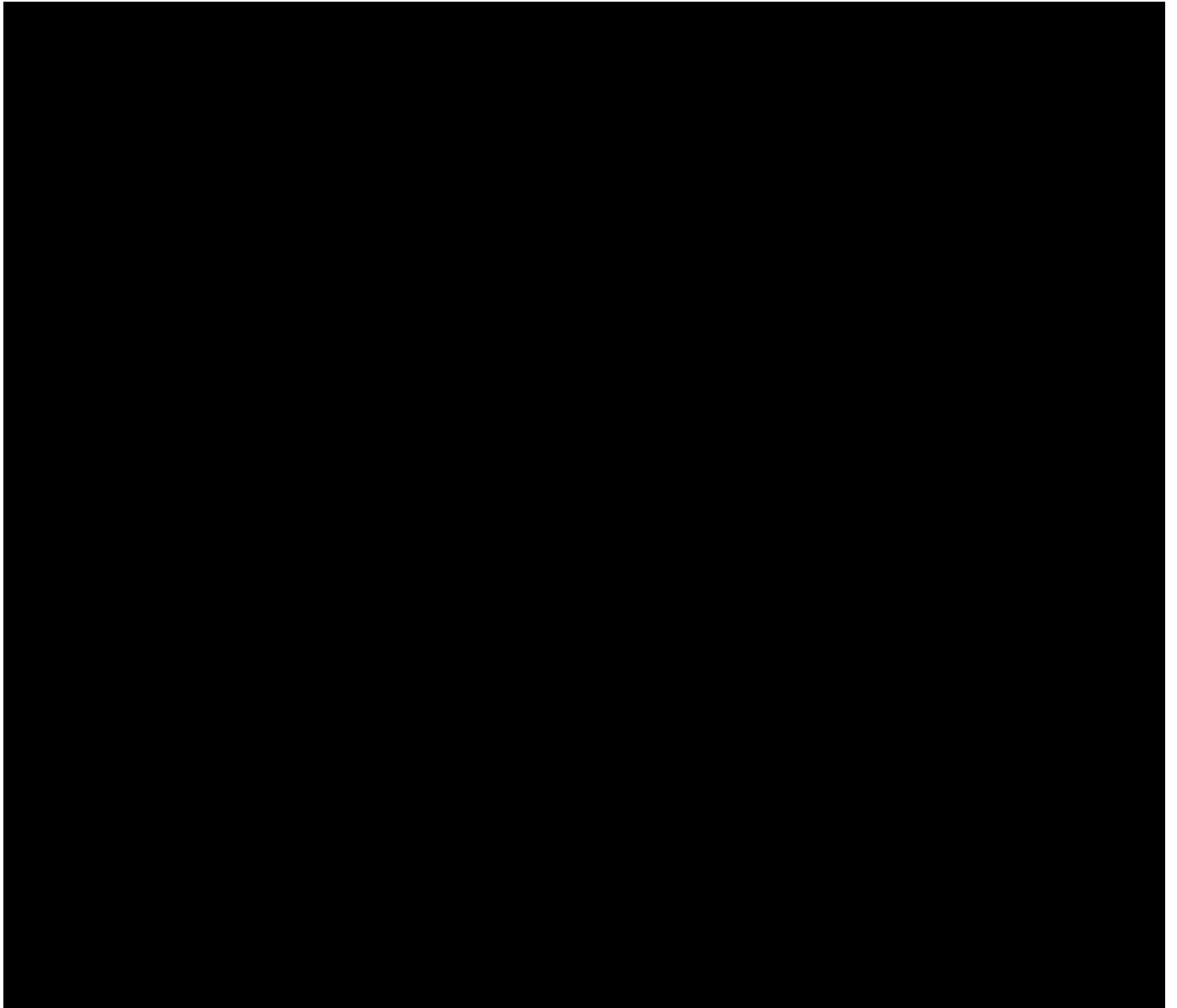
Table 5. Impact of School Racial Composition on Latino Youth Perception of Injustice

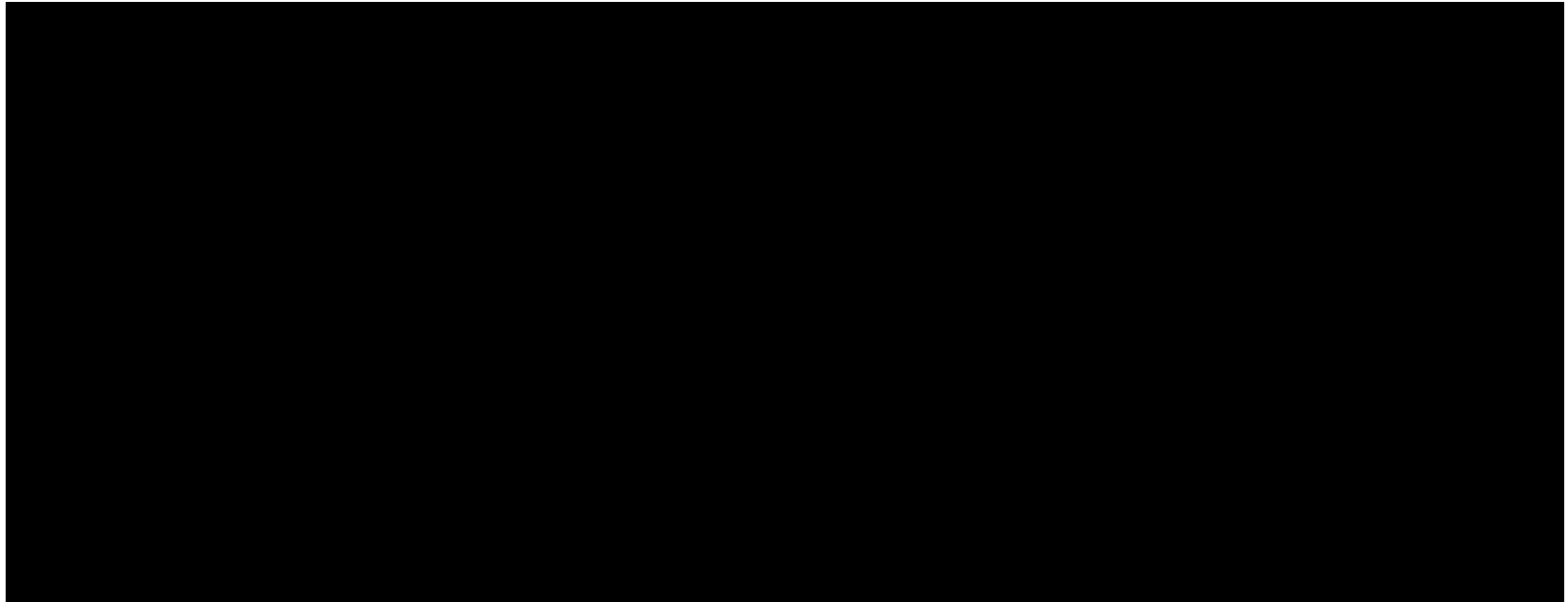Table 3.   Racial/Ethnic Disparities in Arrest, With Individual and Neighborhood Characteristics

Table 4. Racial/Ethnic Disparities in Arrest, With Individual, Family, and Neighborhood Characteristics
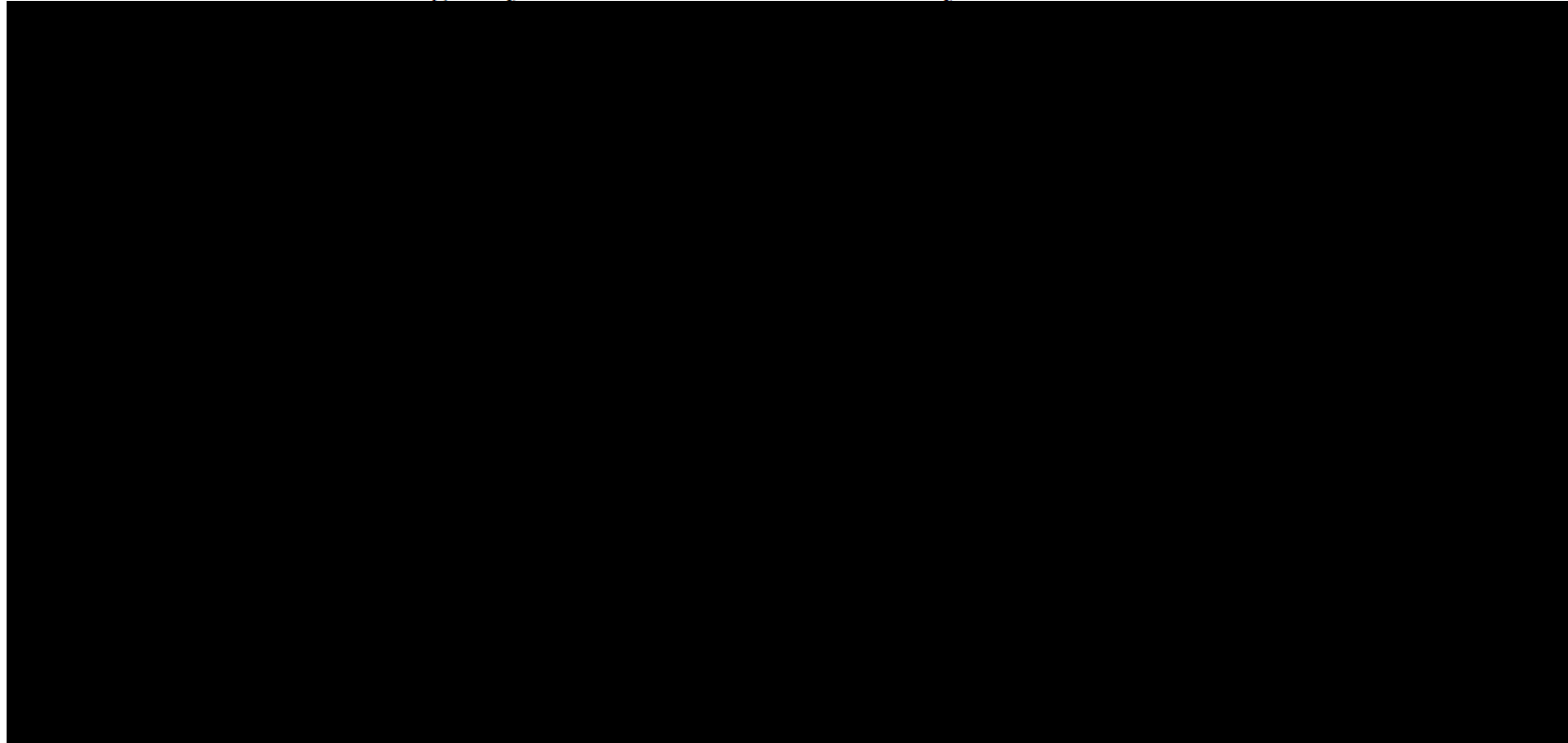
Table 2. Multilevel Model of Arrest with Individual, Family, and Neighborhood Characteristics
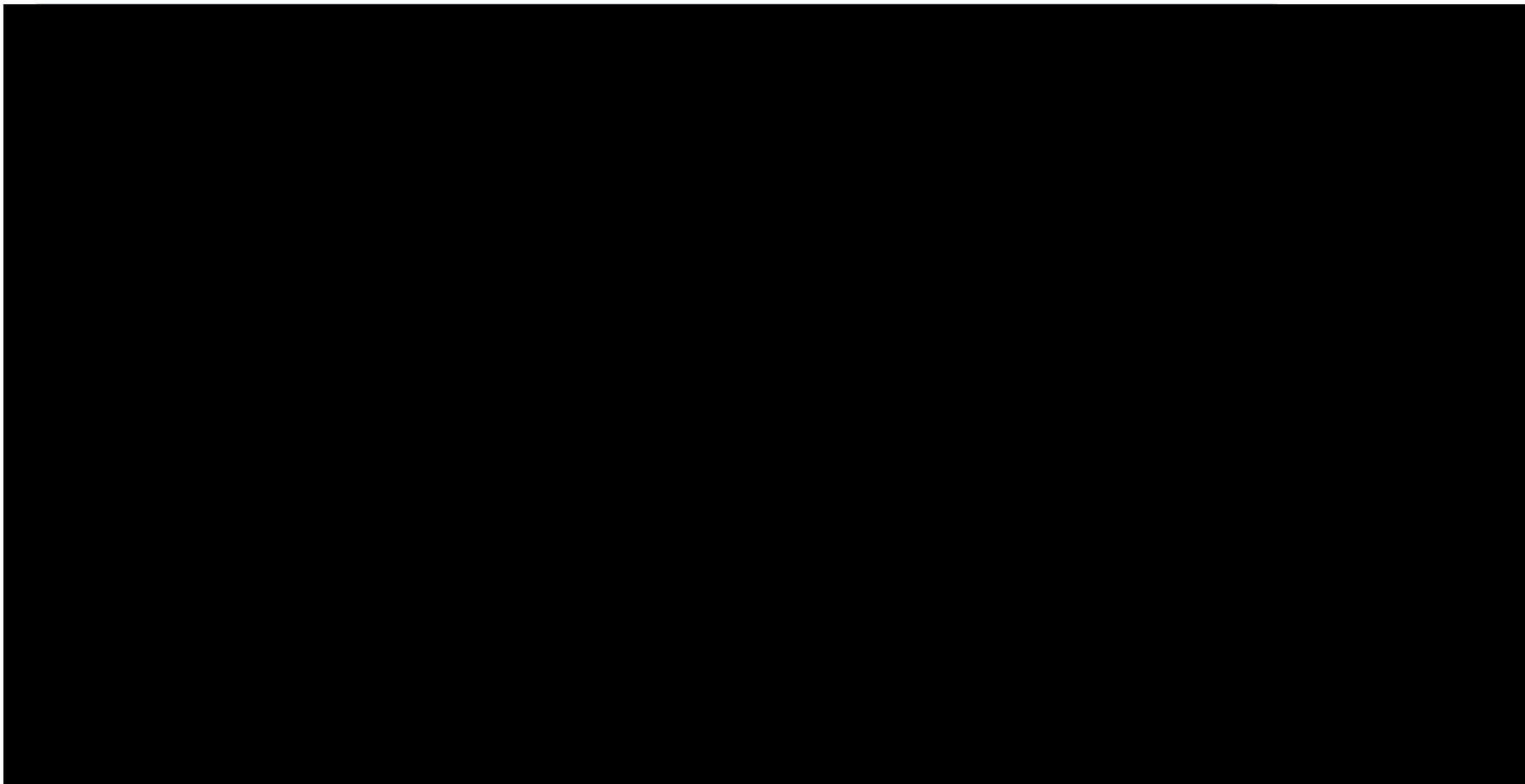
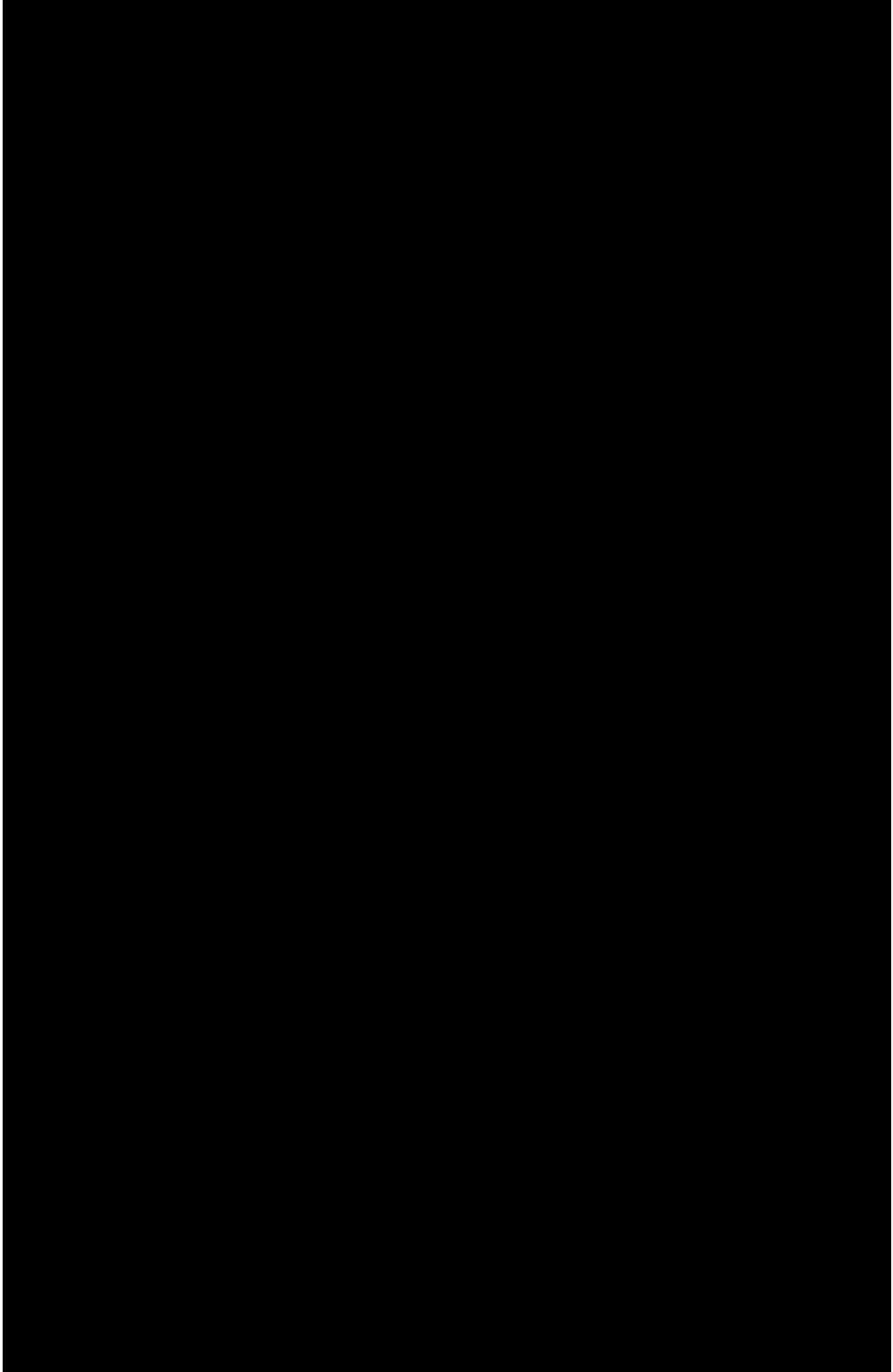**Table 3. The Effect of Legal Cynicism and Collective Efficacy on Arrest**

TABLE 3—Neighborhood Predictors of Violence: Project on Human Development in Chicago Neighborhoods Waves 1–3, Age Cohorts 9–18[a,b]
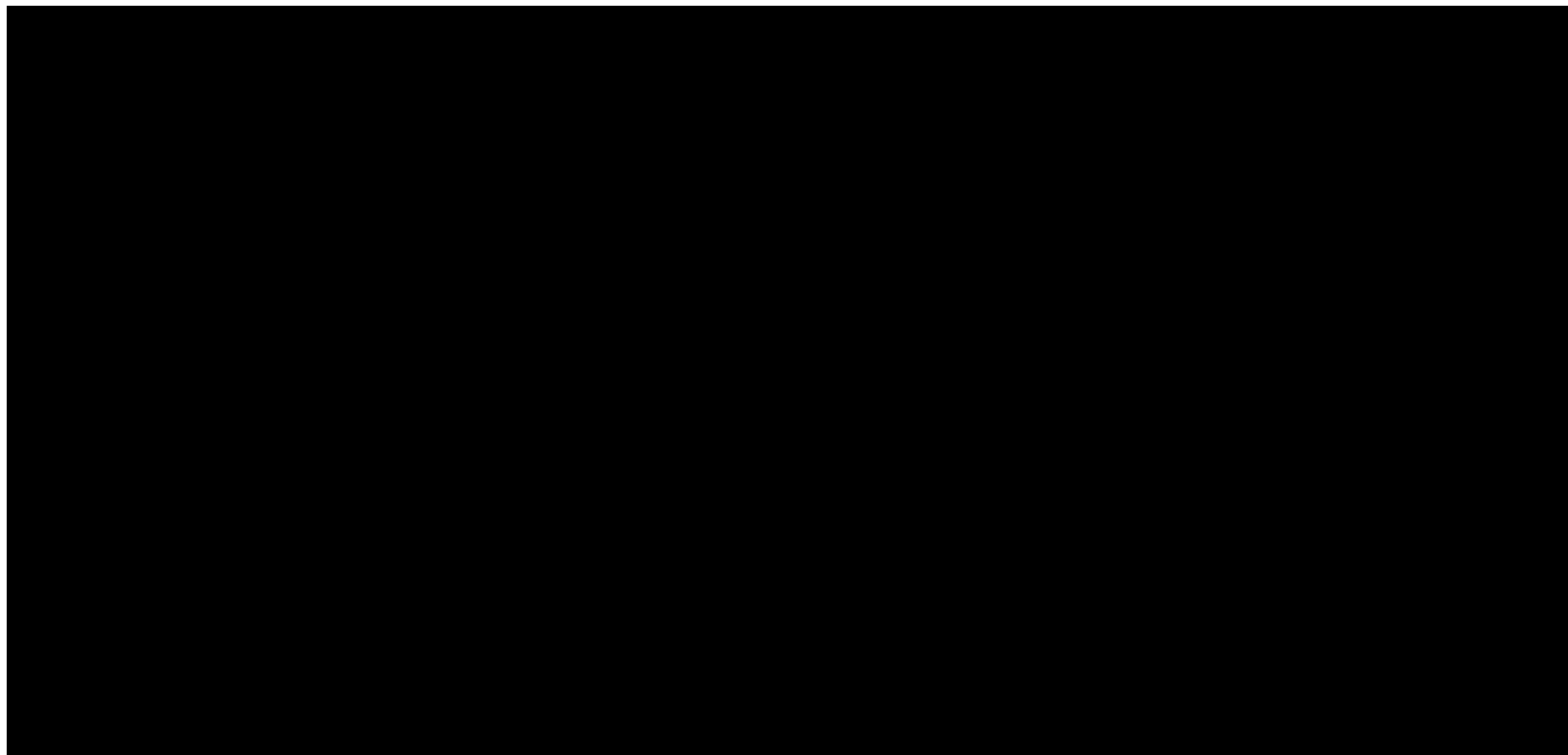
Table 2. **Hierarchical Logistic Regression Models of Female Victimization by Intimate Partner Violence in 40 MSAs, 1989–2004 (*N* = 487,166)**

Table 2. Continued

Peer Panel Document

Re: Eric Stewart

#9

TO: Gary K. Ostrander, Deciding Official
Vice President for Research

FROM: Diana Key, Research Integrity Officer (RIO)   7/30/19
Director, Research Compliance Programs

Attached is the final report from the Inquiry Committee. Please review this document and, as Deciding Official, determine next steps in this case. Please mark your response below as appropriate.

☒ I <u>accept</u> the findings and recommendations of the Inquiry Committee, and direct the RIO to proceed in accordance with FSU policy and procedures.

☐ I <u>do not accept</u> the findings and recommendations of the Inquiry Committee. I direct the RIO to reconvene the Inquiry Committee for further fact-finding and analysis as follows:

_____

_____

_____

_____

_____

_____

_____

☐ I <u>override</u> the findings and recommendations of the Inquiry Committee with the following determination/decision:

_____

_____

_____

_____

_____

_____

_____

Deciding Official Signature/Date: _Gary K. Ostrander_   8/1/19

CONFIDENTIAL

Florida State University
Inquiry Report Concerning
Allegations of Research Misconduct against
Eric Stewart

1.   **Name and position of the Respondents**

Eric Stewart, Professor, Criminology

2.   **Description of the allegations of research misconduct**

Please see the allegation dated 05/30/2019 and the documents provided by Justin Pickett, State University of New York, who was a co-author on one of the papers in question.

3.   **The external support pertinent to the allegation**

The National Science Foundation supported the paper labeled "Stewart et al 2015." No funding sources were disclosed in the other papers in question.

4.   **The names and titles of the committee members and experts who conducted the inquiry**

Karin Brewster, Professor, Sociology;

Sonja Siennick, Professor, Criminology;

William Bales, Professor, Criminology;

5.   **Summary of the inquiry process used**

The committee members reviewed the materials provided, interviewed the respondent, and then met as a committee and discussed the evidence and findings to form a consensus recommendation.

6.   **List of the research records reviewed**

Articles named in allegations:

1. Johnson BD, Stewart EA, Pickett J, Gertz M. 2011. "Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment." Criminology, 49(2), 401-441. REF Johnson, et al. (2011).

2. Stewart EA., Martinez R, Baumer EP, Gertz M. 2015. "The social context of latino threat and punitive Latino sentiment." Social Problems, 62(1), 69-92. REF Stewart, et al. (2015). Funding source NSF grant #1023337 to Stewart/FSU and 1023333 to Martinez/NU.

3. Mears DP., Stewart EA, Warren, PY, Craig MO, Arnio AN. 2019. "A legacy of lynchings: Perceived black criminal threat among whites." Law & Society Review, 53(2), 487-517.

4. Stewart EA., Johnson BD, Warren PY, Rosario JL, Hughes C. 2019. "The social context of criminal threat, victim race, and punitive black and latino sentiment." Social Problems, 66(2), 194-221. REF Stewart, et al. (2019).

5. Stewart EA, Mears DP, Warren PY, Baumer EP, Arnio AN. 2018. "Lynchings, racial threat, and whites' punitive views toward blacks." Criminology, 56(3), 455-480. (2018).

CONFIDENTIAL

Also sent with the Allegations was a list of articles referenced by year the survey was conducted as listed below:

2008 Survey

1. Johnson BD, Stewart EA, Pickett J, Gertz M. 2011. "Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment." Criminology, 49(2), 401-441. REF Johnson, et al. (2011). Funding source not disclosed.

2. Stewart EA., Martinez R, Baumer EP, Gertz M. 2015. "The social context of latino threat and punitive Latino sentiment." Social Problems, 62(1), 69-92. REF Stewart, et al. (2015). Funding source NSF grant #1023337 to Stewart/FSU and 1023333 to Martinez/NU.

2013 Survey

1. Mears DP., Stewart EA, Warren, PY, Craig MO, Arnio AN. 2019. "A legacy of lynchings: Perceived black criminal threat among whites." Law & Society Review, 53(2), 487-517. REF Mears, et al. (2019). Funding source not disclosed.

2. Stewart EA., Johnson BD, Warren PY, Rosario JL, Hughes C. 2019. "The social context of criminal threat, victim race, and punitive black and latino sentiment." Social Problems, 66(2), 194-221. REF Stewart, et al. (2011). Funding source not disclosed.

3. Stewart EA, Mears DP, Warren PY, Baumer EP, Arnio AN. 2018. "Lynchings, racial threat, and whites' punitive views toward blacks." Criminology, 56(3), 455-480. REF Stewart, et al. (2018). Funding source not disclosed.

Data files and documents provided by Justin Pickett:

1. Excel file titled "Justin_voting_data" which contains 1,000 rows of data and six columns.

2. Excel file titled "Voting - Data".

3. Emails to FSU and a report regarding his request for the retraction of one of the articles named in the allegation.

Information provided by respondent:

1. Emails between Eric Stewart and the survey data provider Jake Pratton which included datasets from the 2008 survey by The Research Network and datasets from the 2013 survey (January 9, 2008, July 25, 2008, and May 8, 2017)

2. Stata output showing different samples from the Research Network.

3. Emails between Eric Stewart and journal editors.

4. Emails from Eric Stewart relating to the data issues raised.

5. Output from the statistical programs STATA and SPSS that contains the results from statistical analyses conducted by Eric Stewart.

6. Journal and newsletter articles related to the allegation.

   a. Morenoff et al. 2001, Criminology (Published article addressing stability)

   b. Pickett et al. 2012, Criminology (Published article addressing stability)

   c. Rodriguez 2007, Justice Quarterly (Published article addressing stability)

d. Morris and Perry 2016, Social Problems

e. Diekmann 2007, Journal of Applied Statistics (cited by John Smith - not enough zeros in 3rd decimal place)

f. Diekmann and Jann 2011, German Economic Review (critique of how scholars are uing Benford's law and 3rd digit issue)

g. Stewart et al., correction, Criminology

h. Medium.com, "So, You Want to Blow the Whistle"

i. APS Observer, "A Call to Change Science's Culture of Shaming"

j. Bartlett 2018, Chronicle of Higher Education

k. Sciencemag.org, "Meet the 'Data Thugs' Out to Expose Shoddy and Questionable Research"

7.   **Summaries of Respondent Interview(s)**

Each of the three committee members was afforded the opportunity to ask the respondent any questions they wished answered to and any of the committee members could ask followup questions. This process took quite some time and the respondent was very thorough in his answers to the initial questions and to any follow-up questions or requests for clarification of his answers.

The committee asked broad questions about the processes used to generate the data, analyses, and tables, and specific questions about the issues raised in the allegation. The respondent explained his role in producing and reporting the research results, the status of his ongoing examination of the issues raised in the allegation, and the details of his correspondence with the people bringing the allegations. To summarize, the respondent reported that the inconsistencies identified in the allegation stemmed from a complex rolling data collection strategy, errors in transposing the research results, and the complainant's misuse of posthoc techniques for validating research results.

8.   **Committee recommendation and the basis for the recommendation**

Following the interview, the committee members deliberated for some time concerning the respondent's answers and agreed to reconvene after reviewing materials provided by the respondent (as noted in item 6, Information from Respondent.) This second meeting was held on Friday, July 19, 2019. Additional meetings were held on Wednesday, July 24, 2019 (by phone) and on Monday, July 29, 2019.

Based on our reviews of the published work and the supplemental materials as well as our interview of Dr. Stewart, we believe that no further investigation is necessary. We found no evidence that the data used by Dr. Stewart in the five papers at issue were fabricated. We did find evidence that Dr. Stewart incorrectly described the data used in the 2011 paper on which Dr. Pickett was a coauthor. In addition, we found errors resulting from insufficient care in recording research results. Below, we list our findings for each specific allegation.

1.   Anomalies in standard errors and coefficients:

a. Standard errors are stable across models: Respondent provided output from ongoing reanalysis of the dataset in question and committee members observed that many of the standard errors in these models were stable to fourth decimal place. Although this degree of stability may seem unusual, it likely reflects the lack of shared variation between model controls

(i.e., the variables that appeared in all models) and the predictors stepped (singly) into the models.

b. Lack of third-place zeroes: Committee did not have access to output from original analyses so we could not determine whether the original standard errors lacked ending zeros. We did have a copy of an email sent by the respondent to Dean Thomas Blomberg (5-28-2019) addressing this matter. In it the respondent states that he alters coefficients and standard errors that end with zero or have multiple zeros to the right of the decimal so as to minimize presumed concerns of reviewers. The committee members urge the respondent to either report results "as is" or to rescale variables likely to produce exceedingly small coefficients and/or standard errors.

2.   Discordant means and standard deviations: These were explained by the respondent as errors made in transcribing results from output to manuscript tables.

3.   Incongruent changes in Mears et al. (2019): Respondent provided the output that informed these changes and we are satisfied that the published tables accurately represent the analyses.

4.   Identical descriptive statistics across different samples and samples of different sizes: This problem reflects the respondent's method for constructing tables: For ease of formatting, he "overwrites" a table from a previous project. During his interview with the committee, respondent admitted to a lack of care and a failure to proofread.

5.   Unusual changes in sample sizes over time:

a. Pickett's allegation concerning the 2011 paper: An increase in sample size between a conference presentation (N = 868) and published manuscript (1,184). In his review of his own records and data for this paper, Dr. Pickett found that 500 observations were duplicated—but this duplication was inadequate to explain sample size increase. During the interview, the respondent explained the change in sample size as reflecting his addition to the sample of new observations sent by the Research Network, a local polling firm. He also acknowledged the duplication as an error in merging the new observations into the original data, and noted that he has contacted the journal's editor to expect a reanalysis using the corrected data.

b. "John Smith" allegations: A larger sample in the 2015 paper than in the 2011 paper, although both papers used the same data. In addition, as noted by a committee member during the committee's interview with the respondent, the 2015 sample should be smaller, because it was based on non-Latino Whites but the 2011 sample included Black and Latino respondents as well as the non-Latino Whites. The respondent explained the Smith allegations and the committee member's observations as a reflection of the "rolling data collection" procedures used by the Research Network.

The respondent provided two sets of materials, some of which speak to this issue (one set received during the interview, and more data documentation on 7-25-19 in response to a committee request made 7-24-19). The material received on 7-8-2019 shows that the respondent received three data files from the Research Network with a total of 1,372 observations (500 + 425 + 447), approximately the size of the complete data set (1,379) noted in the 2015 publication. The material received on 7-25-2019 describes three data files that the respondent is using to re-do the analyses for the 2011 paper, with a total of 1,007 observations; the respondent did not share with the committee how he created these files.

6.   Incorrect statistics and distributions: The respondent acknowledged errors in the estimation of the median and his reporting of the cumulative income distribution for the 2011

and 2015 papers. These errors, combined with the respondent's reliance on "overwriting" tables (see item 4, above), produced the problems noted by "John Smith."

7.   Improbable survey design and data structure:

a. Cell phone area codes are not a good representation of an individual's place of residence: During the interview, the respondent noted that zip codes, not area codes, were used to identify place of residence. This information was then aggregated to the county level.

b. Source of the 2013 data: The 2013 survey was conducted by students, under the guidance of personnel who had been employed by the Research Network. The respondent's receipt of data was confirmed by his email records.

c. Data were provided to the respondent through personal contacts with Research Network personnel. No external funding was used for data collection.

8.   Number of counties and creation of the "county clusters": In his emails to Diana Key (6-10-2019 and 7-23-2019), Dr. Pickett noted that the 2011 paper reported using data for 91 counties, but respondents in the Research Network data actually represented 292 counties. In discussing this issue during his interview with the committee, Dr. Stewart acknowledged that, although the paper referred to counties and reported descriptive statistics for counties, in performing the analyses, he aggregated the counties into clusters. He provided no explanation for this decision, but committee members note that aggregating the counties allowed Dr. Stewart to use what was then an emerging statistical approach that could not have been supported by the county-level data. Nonetheless, the use of clustering may explain the discrepancy between the sample size presented in the paper and the sample size observed in the original data. This should have been described more clearly in the paper.

9.   Failure to report handling of missing data: During the interview, Dr. Stewart noted that he relied on multiple imputation to handle missing data, yet none of the five papers addressed in our inquiry discuss imputation specifically or the handling of missing data more generally. Methods to resolve missing data have implications for results and it is standard practice for researchers to discuss missing data, even when they use a complete case approach (i.e., dropping observations with missing data on any variable).

9.   **Respondent comments on the draft report**

☐ are attached, **or** ☒ the respondent chose not to provide any comments.

10.   **Whether any other actions should be taken if an investigation is not recommended**
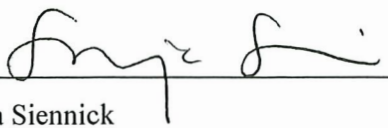
The committee recommends that Dr. Stewart contact the editors of Criminology (2011 paper) and Social Problems (2015 paper) regarding his original description of the "county clusters" as counties and his failure to report his reliance on imputation to handle missing data. The journals' editorial boards will need to decide what steps are necessary and we expect that Dr. Stewart will provide any assistance or information deemed necessary by the journals.

Report submitted by (name and signatures of all Committee members)

William Bales

(signature)

Sonja Siennick

See Addandum about
Karin Brewster

Peer Panel Document

Re: Eric Stewart

#10

CONFIDENTIAL

Florida State University
Inquiry Report Concerning
Allegations of Research Misconduct against
Eric Stewart

1. **Name and position of the Respondents**

   Eric Stewart, Professor, Criminology

2. **Description of the allegations of research misconduct**

   Fabrication/falsification of data included in a publication entitled. "School social bonds, school climate, and school misbehavior: A multilevel analysis." Stewart, Eric A. 2003, Justice Quarterly 20:575-604.

3. **The external support pertinent to the allegation**

   None

4. **The names and titles of the committee members and experts who conducted the inquiry**

   • Kathryn Tillman, Professor, Sociology

   • Fred Huffer, Professor, Statistics

   • Debajyoti Sinha, Professor, Statistics

5. **Summary of the inquiry process used**

   The committee members reviewed the materials provided, interviewed the respondent, and then met as a committee and discussed the evidence, analyzed data using code provided by the Respondent, and formed a consensus recommendation.

6. **List of the research records reviewed**

   Pickett email allegation dated 1/15/2020, with attachments (1) the publication entitled. "School social bonds, school climate, and school misbehavior: A multilevel analysis." Stewart, Eric A. 2003, Justice Quarterly 20:575-604; (2) Appendix A, Comparison to Stewart's (2003: 602); Code for accusations; (3) ICPSR Student data file; (4) Pickett Additional Evidence dated 1/14/2020; (5) Hoffman (2003) Analyses of NLES; (5) Hoffman (2006) Another Analysis of NLES; (6) Peugh (2010) NELS Analysis.

7. **Summaries of Respondent Interview(s)**

   Our committee met with Dr. Stewart on February 11 for about an hour. At this meeting he distributed a document he had prepared which discusses the points raised in Dr. Pickett's email of January 14 attacking the validity of his 2003 paper. This document describes a plausible recreation of a sample similar to that used in his 2003 paper. In our meeting, we questioned Dr. Stewart about the steps in the construction of this sample and the results obtained using this sample, including the loadings obtained in a factor analysis of the school involvement variables. Dr. Stewart also listed in his document various published research studies using the same data which reported a range of sample sizes and degrees of urbanicity, which Dr. Stewart felt contradicted some of Dr. Pickett's assertions. This material was also discussed in our meeting.

8. **Committee recommendation and the basis for the recommendation**

CONFIDENTIAL

Dr. Stewart's paper of 2003 omits many details of the analysis, and he does not have any surviving research documentation which supplies these details. In particular, Dr. Stewart no longer has his original data files nor any of the code he used for his analysis. This rather limits what our committee can do. It is regrettable the Dr. Stewart's paper omitted these analysis details (which relate mainly to the treatment of missing values and the decision of which schools to include), but this in itself does not constitute evidence of research misconduct, and these sorts of omissions were not uncommon at the time of Dr. Stewart's publication.

Dr. Pickett claims that some of the numbers given in Stewart's 2003 paper are "impossible". Dr. Stewart endeavored to construct a plausible recreation of the sample used in his 2003 analysis which produced numbers similar to those from his 2003 paper, thereby refuting their impossibility. Dr. Stewart believed that he would have dropped all schools with small numbers of respondents, and that he would then have imputed replacements for missing values whenever possible. In his plausible recreation, Dr. Stewart first dropped all students with missing School ID's and all students from schools with fewer than 17 student respondents. His intention was then to (1) retain students who had useful values on at least one of the GPA measures (dropping the others), (2) retain students who had useful values on at least one of the School Involvement measures (dropping the others), and, finally, (3) drop students who were missing any of the School Misbehavior measures. In the course of dropping these students, some schools dropped below 17 students, and he then deleted these schools from the analysis. This left Dr. Stewart with a sample consisting of 12,250 students from 569 schools. These numbers are not too greatly different from the 10,578 students in 528 schools reported in Dr. Stewart's 2003 paper. Dr. Stewart then carried out a factor analysis (using PCA) on the correlation matrix of the school involvement measures and obtained factor loadings which were somewhat smaller but had a similar general magnitude to those reported in his 2003 paper. This demonstrates that these numbers are "possible" (although, as noted below, they are not likely to be statistically valid, and we agree with Dr. Pickett that factor loadings this large are highly implausible).

The committee obtained Dr. Stewart's code and exactly replicated his findings described above. We discovered, however, that in deleting the students in steps (1), (2), (3) above, Dr. Stewart made coding errors so that his final data contained many students with numerical missing value codes which were then used in computing the correlation matrix, resulting in greatly inflated factor loadings. If Dr. Stewart's original code contained similar errors, that would explain the factor loadings in his 2003 paper.

(We note that by varying the details of Dr. Stewart's three steps it is actually possible to get much closer to the figures of 10,578 students in 528 schools than he did in his plausible recreation.)

Dr. Pickett notes other irregularities in Stewart (2003). The sample produced in Dr. Stewart's plausible recreation does not succeed in explaining these irregularities; it does not reproduce the "urbanicity" reported in his 2003 paper, nor does it reproduce the number of students "never having been put on an in-school suspension". However, there are many ways to vary the schools which are used and the treatment of the missing values, so that it is conceivable that there exists some reasonable way to construct the sample which actually leads to the values reported in the 2003 paper. And if one allows for the possibility of coding errors, then almost anything is possible. Dr. Pickett notes that the reported mean and standard deviation of the binary variable "School location" are impossible since these values are connected mathematically. These values can be easily explained either by a typographical error or by a coding error. For example, if Dr. Stewart left missing value codes in his original data (as he did in

CONFIDENTIAL

his plausible recreation) and used these in his calculation of the mean and SD , then the mean and SD are no longer mathematically related in the expected way.

Our committee has found no evidence which strongly points to research misconduct. Given the many conceivable ways the original analysis could have been conducted and the possibility of coding errors, it does not seem feasible to reconstruct the original analysis or even determine with any certainty for many quantities what values are possible or would indicate research misconduct. For these reasons we believe that a full investigation is not warranted.

9.   Respondent comments on the draft report
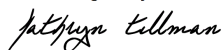
☐ **are attached, or** ☒ **the respondent chose not to provide any comments.**

10.   Whether any other actions should be taken if an investigation is not recommended

N/A

Report submitted by (name and signatures of all Committee members)
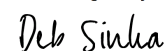
DocuSigned by:

Kathryn Tillman

6E59FA90BC5C4FE...

Kathryn Tillman, Professor, Sociology

DocuSigned by:

Deb Sinha

CBAE0393DAC54A6...

Debajyoti Sinha, Professor, Statistics

DocuSigned by:

Fred Huffer

1843959E285E45F...

Fred Huffer, Professor, Statistics

Peer Panel Document

Re: Eric Stewart

#11

MEMORANDUM

Date: April 1, 2021

To: Dr. Gary Ostrander

From: Eric Stewart

Dear Dr. Ostrander:

I am writing to provide a theoretical and methodological justification for each of the five retracted articles published in criminological and sociological journals. The discussion proceeds as follows. First, I begin with an Introduction that explains the veracity and authenticity of the data utilized in each of the articles. This includes providing written documentation from the owner of the Research Network who documents the numerous datasets provided to me from 2008 through 2018. I then provide a discussion of the methodological decisions made across each article. Following these discussions, I outline each article as well as the articles' results. Additionally, in Appendix A, I have provided a table which lists a sample of published articles that utilized the data collected by the Research Network and have an over-presentation of Whites, middle-class, college educated, and individuals in the southern region. Therefore, the data presented in the retracted articles are consistent with other data produced by the Research Network. In summary, this memorandum provides detailed information reinforcing the inaccuracy of the research misconduct allegations.

Sincerely,

Eric A. Stewart

Eric A. Stewart

Introduction

The following discussion provides a theoretical and methodological explanation of the five retracted articles. Prior to my discussion of the articles, I would first like to explain the veracity and authenticity of the data utilized in each of the articles. This memo substantiates the fact that the published works are not based on fabricated or falsified data or analyses. The data utilized for the articles were collected by the Research Network, a survey research center. Dr. Marc Gertz, a faculty member within the College of Criminology & Criminal Justice, owned and operated the Research Network for 26 years. During that time, he would permit me to add sets of questions to numerous surveys being collected by the Research Network to study issues related to race, ethnicity, discrimination, and punitiveness (Note: Dr. Gertz extended this same courtesy to other faculty and graduate students within the College of Criminology). I was gathering data because I was building one of the most comprehensive databases in the field of Criminology to study complex issues around environment, race, ethnicity, discrimination, and punitive sentiment. To accomplish this task, there were multiple surveys in which I was allowed to add questions. There were also other datasets that were provided to me because they asked similar questions related to race, ethnicity, discrimination, and punitiveness and demonstrated high degrees of internal consistency (α = .70 or higher) among the questions. I was consistently provided with datasets in order to build a larger database. Dr. Gertz provided me with datasets in 2008, 2009, 2013, 2017, and 2018. Following this Introduction, I have provided an email from Dr. Gertz that clearly states the years for which data were shared with me. Between 2008 and 2018, I received more than 10 separate datasets from the Research Network. The datasets range in sample size from 450 to 3,000. Each data file contained demographic information including race, ethnicity, gender, age, marital status, educational attainment, income, city or zip code in which the respondent currently resided, and numerous questions relevant to studying issues about race, crime, and justice in America.

After receiving each dataset, I would ███████████████████████████████████████
██
██
██
██
██
██
██
██
██
███████████████████████████████████████. In the manuscripts exploring lynchings, the lynching data for Southern states were provided by Stewart E. Tolnay and E. M. Beck. The data reflect lynchings that occurred from 1877-1950. Importantly, by 2017, I had a data file which included approximately 5,500 survey respondents. In each of the published studies, I would ████████████████████████████████
██
██
██
██
██
██
██

███
██

███
███
███
███
███
███████████████████████████████

For the Stewart et al. (2018, *Criminology*; 2018, *Social Problems*) and Mears et al. (2019, *Law and Society Review*) studies, I relied on public opinion data that were received from the Research Network in 2013. It was incorrectly assumed and reported that the data were collected in 2013. In prior years, I was forwarded data in the year in which the data were collected. However, after the articles were published (and questions raised about the studies), Dr. Gertz clarified that while the data used for the three studies were distributed to me in 2013, these data were collected between January 2008 and December 2010. I notified the editors of the three journals to report this error and offer corrections.

███
███
███
███
███

Importantly, this rounding did not and could not change the substantive or statistical significance of the papers' findings. While applying this rounding rule did not impact the substantive conclusions reported in the articles, I should have explained this process more clearly.

In the articles, the standard deviations for some of the binary measures were incorrect. This is because I mistakenly reported the standard deviations of those variables in their original ordinal scales rather than the recoded binary versions of these measures used in the studies (the different versions of the variables had very similar names, and unfortunately in some instances the wrong versions were included in the descriptive statistics reported in the articles). It is important to note that unlike standard deviations for continuous measures, standard deviations for binary measures are not readily interpretable and do not provide information unique to the sample.

Also of note, in the articles, I used ██
███
███
███
███
███
███
███
███
███
███
███
███
███

Next, the demographic characteristics in all five of the manuscripts are consistent with the samples in other manuscripts utilizing data collected by the Research Network; which were nationally representative samples of the U.S. population. The overrepresentation of several sample characteristics is also reflected in other studies that used Research Network data (see Table 1 in Appendix A). For example, in a paper published in *Criminology* by Chiricos et al. (2000), 73% of the total sample is White, 11% is Black, 16% Hispanic, and respondents are approximately 45 years of age. Similarly, Pickett et al. (2014) has a published manuscript that uses a sub-sample of the entire survey population. The sub-sample is White and represents 80% of the total sampled population. Additionally, similar to Johnson et al. (2011), 58% of the sample in Pickett et al. (2014) is female, 32% is 65 years of age or older, 32% graduated from college, and 34% of the sampled population is in the Southern region. In another paper published in *Criminology*, Chiricos et al. (2004), report 79.8% of the total surveyed population is White, 11.4% is Black, and 7.4% is Hispanic. Approximately, 44% graduated from college with a mean age of 46. Finally, in a manuscript published in *Crime and Delinquency* by Stupi et al. (2016), also using Research Network data, the sample is 84% White, 9% Black, with an average age of 43 and 55% of the sample residing in households with incomes of at least $50,000.

Across 16 years, the overrepresentation of Whites, females, older respondents, college educated respondents, and respondents from the southern region observed in my published work is reflective in other published manuscripts that utilized data from the Research Network also appearing in *Criminology, Justice Quarterly,* and other top journals in both sociology and criminology. Moreover, despite the various sample sizes across the studies listed in Table 1, the demographic characteristics are incredibly consistent and demonstrate a broader pattern that is reflective of the demographics in the retracted articles being questioned in this context.

In sum, the purpose of this narrative is to clarify decisions that were made relative to the data and results published in the five articles that were retracted. As I outlined above, some of the concerns about the manuscripts were a function of unfortunate coding and transcription errors, along with insufficient information provided in the manuscripts about the data and methodological decisions. While I regret these errors, there was no bearing on the substantive conclusions drawn. In retrospect, it is clear that a more detailed explanation of the data and/or methodological decisions could have been instrumental in heading off some of the concerns. As a result of the errors observed in the manuscripts, my co-authors and I decided to voluntarily retract the manuscripts because we were uncomfortable with the number of errors in the published manuscripts. Despite malicious allegations of research misconduct grounded in highly inflammatory language, there was no fabrication or falsification of data or results in any papers that I have authored. The unconscionable act of publicly accusing scholars of fraud without proper investigation, full knowledge, and correct information creates an atmosphere in which errors in research are assumed to reflect fraud rather than unintended and correctable mistakes that occur in research. In fact, corrections of major analytical errors have been issued in social science journals in recent years without accompanying allegations of research misconduct. The violation of confidentiality by the public sharing of partial and inaccurate information amplified this situation to its current standing. The public and incendiary nature of the misconduct accusations created a highly toxic environment where there was an inherent willingness and urgency to discredit my record.

References:
Royston P. 2005a. Multiple imputation of missing values: update. Stata Journal 5:188–201.
Royston P. 2005b. Multiple imputation of missing values: update of ice. Stata Journal 5:527–536.
Rubin D.B. 1987. Multiple imputation for nonresponse in surveys. Wiley, New York.

From: Marc Gertz mgertz@fsu.edu
Subject: Letter
Date: Aug 6, 2019 at 3:28:44 PM
To: Eric Stewart estewart2@fsu.edu

To Whom it May Concern,

My name is Marc Gertz and I was the founder and president of The Research Network (a survey research firm) for 26 years. During that time I conducted myriad academic research projects for numerous colleagues, including many for the College of Criminology at FSU, as well as other departments at FSU and other universities. A few surveys were conducted in Europe and the Middle East. Many of these data sets ended up as scholarly articles in top tier journals and doctoral dissertations.

clearly incorrect.
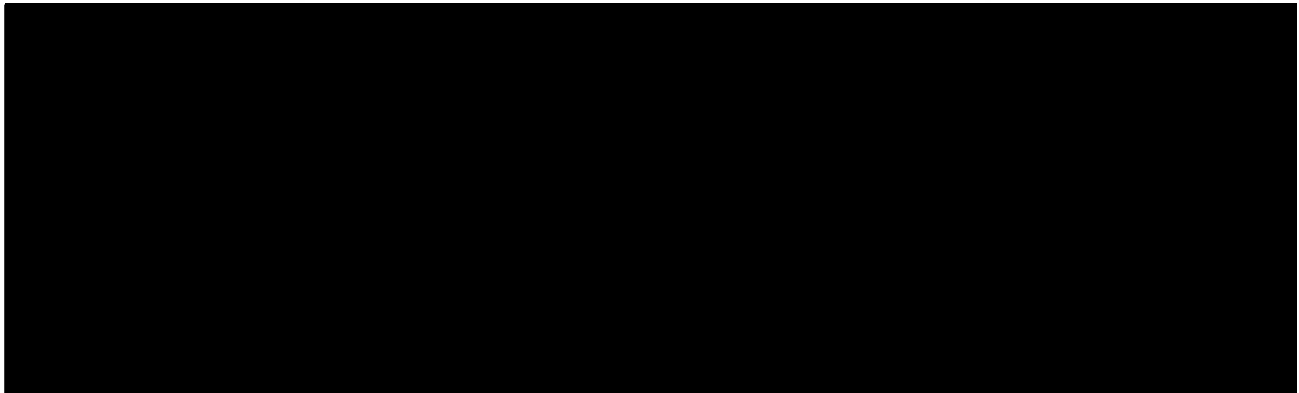
If you have further inquiries, please do not hesitate to call on me.

Thank you for your time and attention
Sincerely, marc gertz

Ethnic Threat and Social Control: Examining Public Support for Judicial Use of Ethnicity in Punishment. (2011).

Purpose of the Study

This research is one of the first attempts to investigate the role that both objective and perceptual measures of ethnic threat play in the formation of public support for judicial use of offender ethnicity in sentencing. Thus, the study responded to calls for additional research that explored multiple dimensions of group threat, for additional minority groups, and for additional measures of social control. This study incorporated both objective and perceptual threat measures to disentangle their unique and complementary influences. To investigate these issues, the study specifically examined six research hypotheses:

H1: The relative size of the Hispanic population will be positively related to public support for use of ethnicity in criminal sentencing.

H2: The growth rate of the Hispanic population will be positively related to public support for use of ethnicity in criminal sentencing.

H3: Perceptions of ethnic criminal threat will be positively related to public support for judicial use of ethnicity in criminal sentencing.

H4: Perceptions of ethnic economic threat will be positively related to public support for judicial use of ethnicity in criminal sentencing.

H5: Perceptions of ethnic political threat will be positively related to public support for judicial use of ethnicity in criminal sentencing.

H6: The effects of perceived political, economic, and criminal threat on public support for judicial use of ethnicity in criminal sentencing will be greatest in social contexts characterized by relatively large or increasing Hispanic populations.

███
██
██████████████████████████████

      ██
██████████████

      ██
████████████████████

      ███
██████████████████████████

      ███
████████████████████████████████████

█████████████
██
██
██
██
██

█████████████
██
██
██
██
██
██

Results

1. The results suggested that changing population demographics were a strong predictor of support for judicial use of offender ethnicity in sentencing. Specifically, the growth rate of the Hispanic population was strongly and positively related to support for use of ethnicity in sentencing.

2. The results provided empirical evidence for the link between criminal threat and approval of the use of ethnic considerations in punishment. Specifically, individuals who perceived Hispanics to be a greater threat to public safety were more likely to support judicial use of ethnicity in sentencing.

3. The results also suggested that individual perceptions of economic threat are positively related to public support for use of ethnicity in the exercise of social control. Individuals who perceived Hispanics to be a greater threat to scarce economic resources like employment and welfare were more likely to support judicial use of ethnicity in punishment.

4. The results showed that perceived threats demonstrated stronger influences in contexts characterized by higher levels of objective threat. Both criminal and economic ethnic threat measures became stronger as the Hispanic growth rate increased in the county.

Description of Measures

Social Context of Latino Threat and Punitive Latino Sentiment. (2015).

Purpose of the Study
The relationship between racial context, extralegal factors, and social control has long been a topic of debate among sociologists and criminologists. Much of this work is rooted in the conflict and racial threat perspectives, which argue that culturally dissimilar minority groups are perceived by the dominant group as a diffuse threat to the social order, and that the criminal justice system is instrumental in helping to control such threats. Thus, using the Latino group threat perspective, this study investigated the effects of micro- and macro-level measures of Latino threat on punitive Latino sentiment. The research was directed at advancing understanding of the role that Latino population size and growth played in shaping individual-level perceptions of Latino threat and attitudes among whites that are consistent with a desire for greater social control efforts against Latinos. The following hypotheses were explored:

H1: Does the relative size of the Latino population predict punitive Latino sentiment among Whites?

H2: Does growth in the Latino population predict punitive Latino sentiment among Whites?

H3: Is criminal, economic and political threat posed by Latinos associated with punitive Latino sentiment?

H4: Are perceptions of Latino threat associated with punitive Latino sentiment?

Results

1. The findings indicated that a growing ethnic minority population increases perceived threat among Whites, which, in turn, heightened support for punitive social control of ethnic minorities. Specifically, the findings suggested that about one-third of the effect of Latino population growth and more than one-half of the effect of the relative size of the Latino population on White punitive sentiment toward Latinos is due to heighted perceptions of Latino economic and criminal threat.

2. The results indicated that dimensions of perceived Latino threat translated into greater punitive sentiment against Latinos particularly in areas with recent growth in the Latino population. These findings imply that Latino population context is an important dimension of shaping White attitudes about perceived economic and criminal threat, which in turn yields greater punitiveness.

Description of Measures

Variable	OPERATIONALIZATION

Lynchings, Racial Threat, and Whites' Punitive Views towards Blacks (2018).

Purpose of the Study

An emerging body of research suggests that past lynchings are associated with contemporary racial disparities in crime, punishment, and social policies. This study sought to contribute to scholarship aimed at testing racial threat theory and advancing understanding of the factors that give rise to punitive sanctioning, racial disparities in punishment and, in turn, racial divides in contemporary America. Specifically, the study sought to understand Whites' views about punishment of Black criminals and the role that lynchings play in shaping these views. There are no direct tests of the implied theoretical mechanisms or, in turn, an interactional model in which the effects of lynchings on Whites' punitive sentiment may be amplified by views of Blacks as criminal threats. Therefore, this research examines the following research hypotheses:

H1: Whites who reside in areas with a higher volume of past lynchings will be more likely to support tougher sanctioning of Black criminals.

H2: The effect of lynching on White punitive sentiment towards Blacks will be partially mediated by Whites' perceived criminal threat from Blacks.

H2a: The effect of lynchings will be partially mediated by perceived Black criminal threat.

H2b: The effect of lynchings will be partially mediated by perception of Black-on-White violence.

H3: The impact of lynching on White punitive sentiment towards Black offenders will be moderated by Whites' perceived threat from Blacks.

H3a: The effect of lynchings will be amplified by Whites' perceptions of Black criminal threat.

H3b: The effect of lynchings will be amplified by Whites' perceptions of Black-on-White violence.
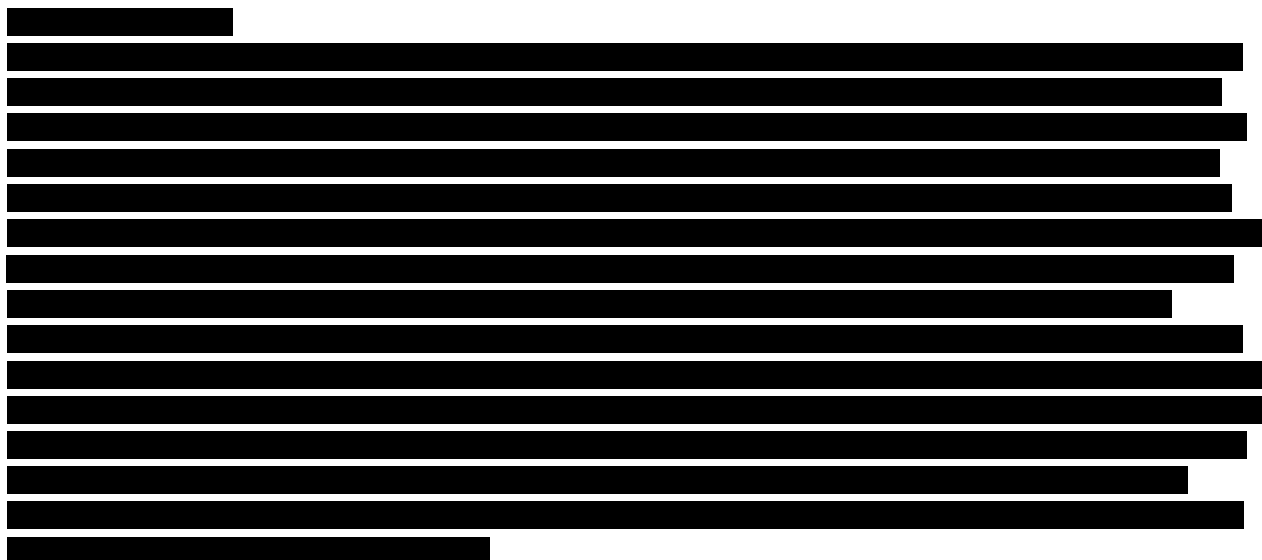
[redacted]

Results

1. The findings indicated that Whites who lived in areas where Black lynchings were more common were more likely to support punitive responses to Black criminals.

2. The results also indicated that Whites supported substantially tougher punishment of Black criminals as compared to White criminals. These effects were partially mediated by perceptions of Blacks as a criminal threat in general and to Whites in particular.

3. The results suggested that when the focus was on Whites' perceptions of Blacks as likely to commit crimes specifically against Whites, an amplification effect emerged. Specifically, among Whites who perceived Blacks as likely to victimize Whites, residing in areas with higher levels of lynchings increased Whites' punitive sentiments toward Black criminals.
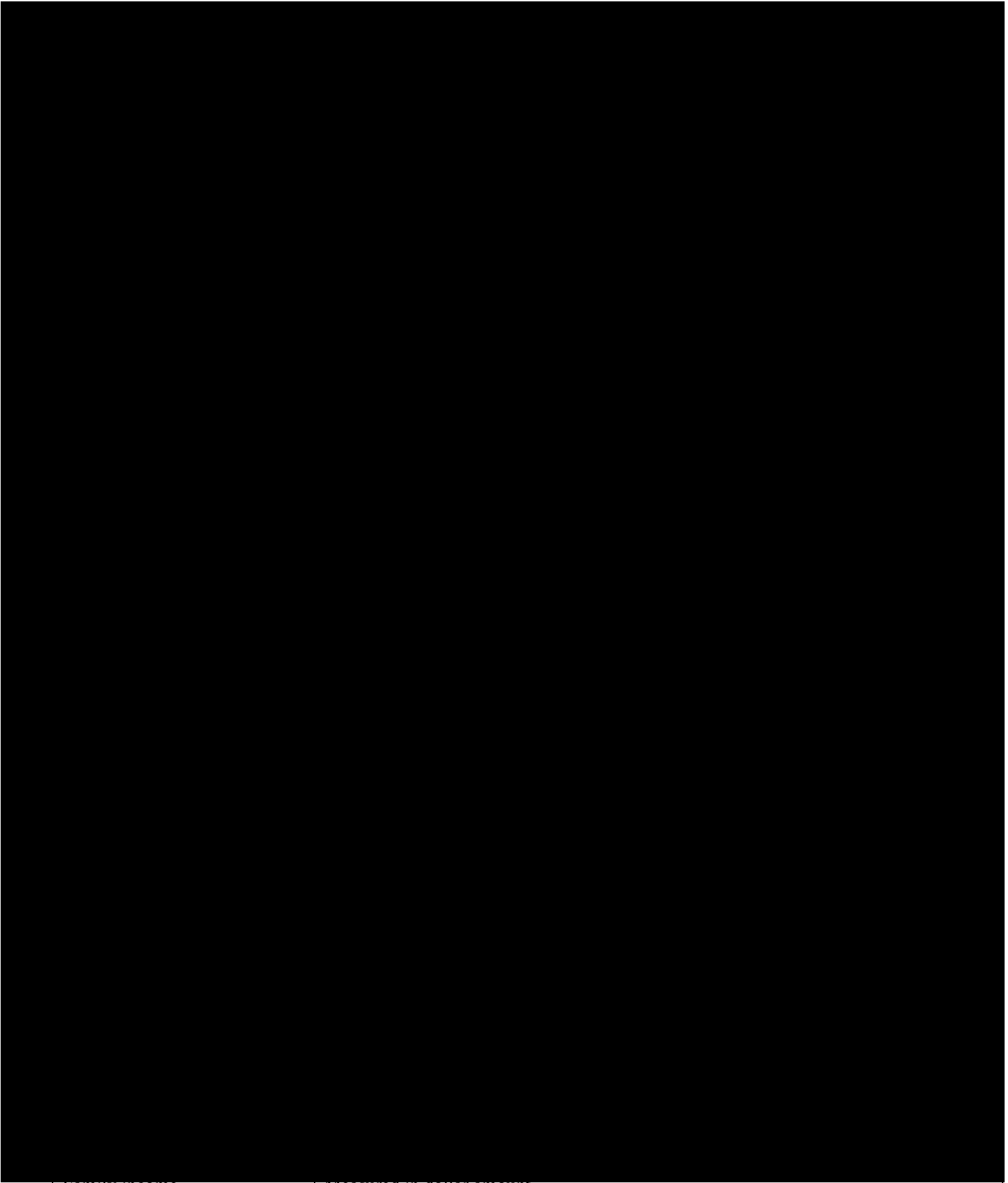
| Family income | Measured in dollar amount |

14

The Social Context of Criminal Threat, Victim Race, and Punitive Black and Latino Sentiment (2019).

Purpose of the Study

Theoretical treatments of race and punishment are routinely cast in terms of racial group threat perspectives that suggest large and growing minority populations generate group prejudice and hostility on the part of the White majority, which is then translated into enhanced social control efforts in the criminal justice system. At the same time, sociological formulations of the behavior of law argue that race is a key component of social stratification in American society and stronger legal responses are expected in situations involving White victims and racial minority offenders. The current research united arguments, examining the complementary and interactive influences of aggregate population contexts, localized perceptions of criminal threat, and race of victim effects in public support for enhanced social control directed at minority offenders. Toward that end, the following hypotheses were explored in this study:

> H1: The objective size and recent growth of the Black and Latino populations will increase punitive sentiment among White respondents.
>
> H2: Subjective perceptions of Black and Latino criminal threat will increase punitive sentiment among White respondents.
>
> H3: Crimes involving White victims that are committed by Black or Latino offenders will increase punitive sentiment among White respondents.
>
> H4: Crimes involving Black or Latino victims that are committed by Black or Latino offenders will not increase punitive sentiment among White respondents.
>
> H5: The effect of victim race on punitive sentiment will be stronger in social contexts characterized by large or growing Black and Latino populations.
>
> H6: The effect of victim race on punitive sentiment will be stronger in social contexts characterized by greater subjective perceptions of Black and Latino criminal threat.
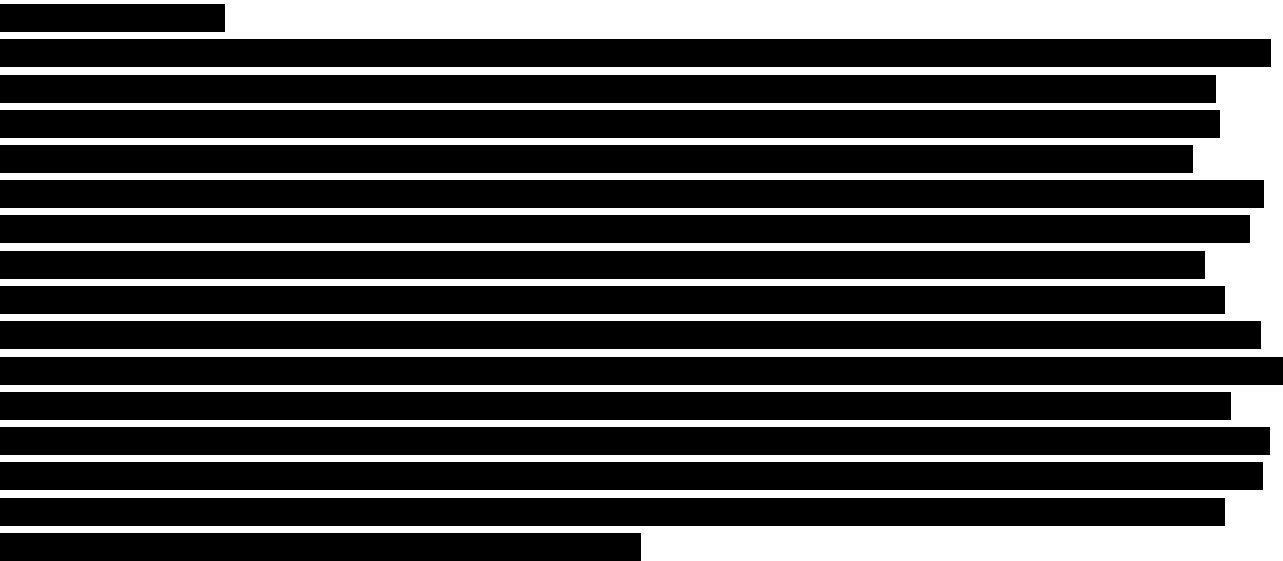
Results

1. The findings showed support among White respondents for punitive sentiment aimed at Black and Latino offenders.

2. The results suggested that punitiveness was significantly related to recent population growth for both groups. White survey respondents in counties with growing Black or Latino populations expressed greater punitive sentiment toward Black and Latino defendants.

3. The findings indicated that perceptions of criminal threat were positively related to punitive sentiment toward Black and Latino offenders. Individuals who viewed Blacks and Latinos as more criminally involved and greater threats to public safety are more likely to support punitive measures that specifically target them in the criminal justice system.

4. The findings suggested that victim race plays a role in punitive sentiment toward Black and Latino offenders. Specifically, White respondents report greater punitiveness when asked about scenarios involving White victims and minority perpetrators. However, for situations involving minority victims, White respondents expressed less punitiveness toward White offenders who target minority victims.

5. The results indicated that the effect of a White victim was stronger in areas where Black criminal threat was high and Black population growth was more rapid. This effect also increased significantly when Latino criminal threat was high and Latino population growth increased.

Description of Measures

■■■	■■
	■■■■■■■■■■■■■■
■■■■■	■■
	■■■■■■■■■■■■
■■■■■■■■■	■■
	■■■■■■■■■■■■■■■■
■■■■■■■■	■■■
	■■
	■■
■■■■■■■■■	■■■
	■■■■■■■■■■
■■■■■■■■■■■	■■■
	■■■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■■■	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■■■■■■
■■■■■	■■■■■
■■■■■	■■■■■
■■	■■■
■■	■■■■■
■■■■■	■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■■■■■■■■■
■■■■■■■■■■■■■	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	■■■■■■■■■■■

A Legacy of Lynchings: Perceived Black Criminal Threat among Whites. **(2019).**

Purpose of the Study

Research on lynchings has emphasized the salience of lynchings for exemplifying and supporting a culture of racial animus and hostility towards Blacks that exerts a persisting influence on race relations in contemporary America. This paper sought to contribute to scholarship aimed at understanding the historical legacy of lynchings and contemporary racialized views of crime. Specifically, it examined whether lynchings influenced modern-day views that Whites hold of Blacks both as criminals and as criminal threats to Whites. This theoretical argument stemmed from literature on the role of lynchings in expressing and supporting a deep-rooted cultural view of Blacks that persists in contemporary society and continues to shape how Whites perceive Blacks. Drawing on this work and on racial threat theory, this research examines the following research hypotheses:

> H1: Perceived Black criminal threat—that is, the perception that Blacks are criminal and that they are more likely to commit crimes against Whites—will be greater among Whites who reside in areas that experienced higher numbers of lynchings.

> H2: In contemporary America, this effect will be more pronounced among Whites who reside in areas marked by greater social and economic disadvantage.

> H2a: In contemporary America, this effect will be more pronounced among Whites who reside in areas that are politically conservative.

Results

1. The results indicated that Black lynchings were significant and positively associated with perceived Black criminal threat and perceived threat of Black-on-White crime. This finding suggested that, among Whites in contemporary America, levels of perceived Black criminal threat and threat of Black-on-White crime are higher in counties where historical patterns of Black lynching activities were more frequent.

2. The results indicated that concentrated disadvantage and percent voting Republican are associated with Whites' views of Black criminal threat and perceived threat of Black-on-White crime.

3. The findings showed that the effects of lynchings on perceived Black criminal threat and perceived threat of Black-on-White crime are stronger in disadvantaged counties and in counties where the percent voting Republican is greater.

Description of Measures

Appendix A
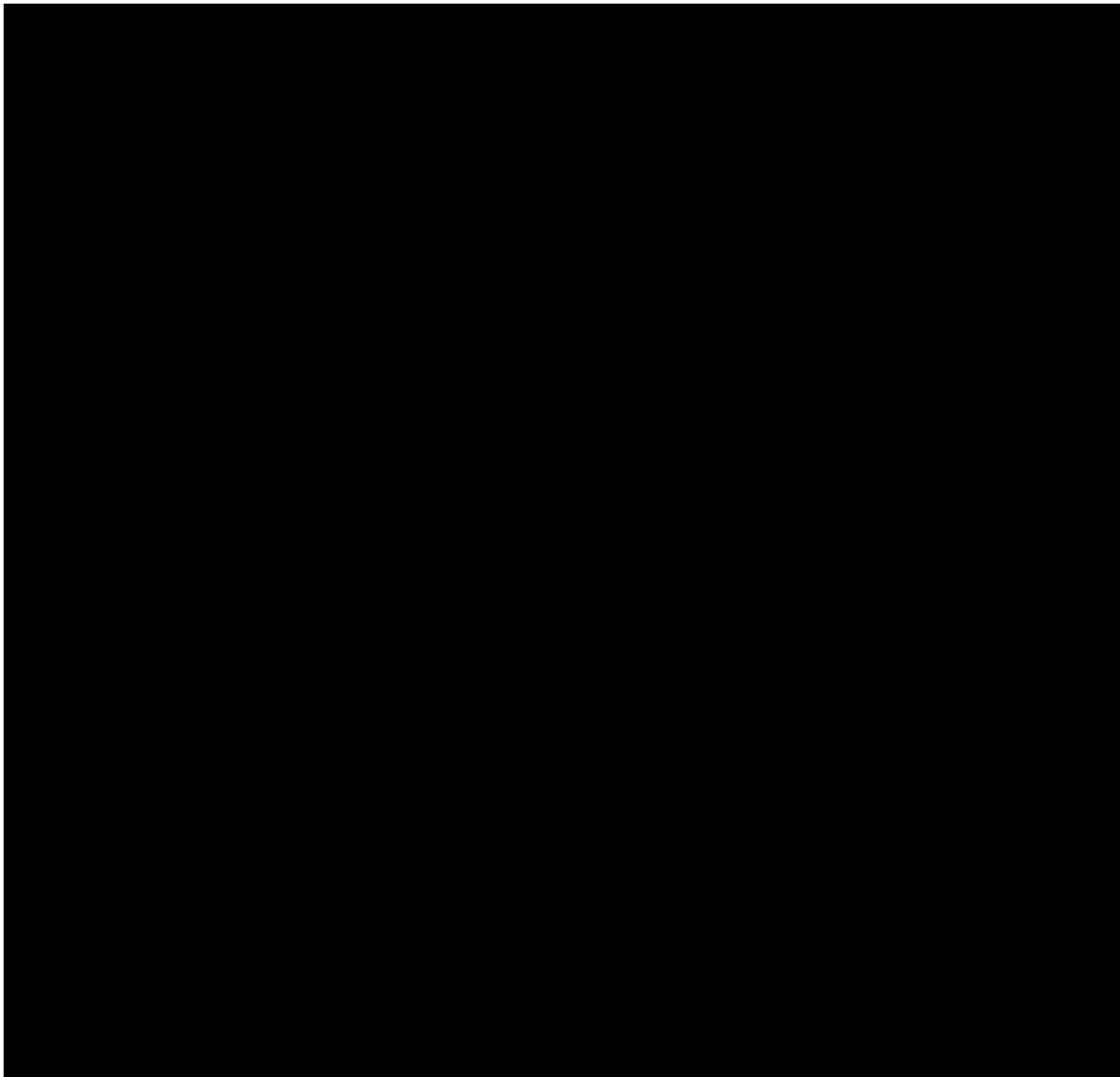
Peer Panel Document

Re: Eric Stewart

#12

FLORIDA STATE UNIVERSITY
OFFICE OF HUMAN RESOURCES

Investigatory Interview: Eric Stewart
Date: April 11, 2022
Attendees:
Rebecca Peterson- Faculty Relations, Associate Director
Eric Stewart, Professor
Thomas Blomberg, Dean, Criminology

What is your role at the University and how long have you been employed?

Professor of Criminology and been employed 15 years.

What courses do you teach? What is your area of research?

Research methods and Statistics, Violence in America, Serial Killers, Victimization, American Homicide, Intro to Criminal Justice, Gangs and Society, Cyber Crime and Victimization, Homeland Security and Terrorism, Private Security, and The Nature of Violence are courses I teach. My research areas include Neighborhoods and Crime, Social Context and criminal justice outcomes, and Crime over a life course, Crime trends. Trying to understand how context (counties, neighborhoods, and schools) relates to criminal outcomes (offending, victimization, arrest, incarceration, punitive attitudes, and suspensions).

What is your training in statistical analysis?

Pretty extensive as far as social science is concerned. My statistical training began with my master's degree and continued post-doctoral degree.

Are you familiar with FSU Policy 7A-26 Research Data Management? What is your understanding of your responsibility as a researcher in data collection and retention?

After a certain amount of time, typically three years, I retain data that long and then I purge the data but maintain data records. If FSU had a policy beyond that, I wasn't familiar with it. In some cases with restricted data that I use, I don't keep data longer than two years past the project period because of the sensitive nature of the data; but I would maintain data records. I typically stored data and data records on my university issued, password-protected laptop computer. Given the varying data management protocols for the multiple sets of data that I worked with, I self-imposed stringent rules for securing the data (i.e., use of a password-protected computer; computer was not connected to a network; data was not stored on a separate storage device; files were not backed up [stored] on a "cloud" server; paper records were purged within the requisite time frame).

What is the industry norm for keeping data analysis information related to publications? How long should data be kept?

Yes, you want to maintain data and records for some period of time. The exact period of time is not clear because it depends on the type of data (restricted versus not restricted) one is using and the procedures set forth by the individual or agency supplying the data.

It was reported that there were concerns regarding several of your publications which prompted an inquiry from the Office of Research. When did reports regarding your analysis or data first surface? What were the concerns?

I believe it was in 2019. It could have been April or so but it was in 2019. The concerns were that some of the binary standard deviations were not correct. The regression coefficients were too stable in our

FLORIDA STATE UNIVERSITY
OFFICE OF HUMAN RESOURCES

multilevel models and there weren't enough zeros in the third column past the decimal place. I took the values out three decimals places. But there weren't enough zeros in the third column—I don't know what that means. The argument is that there should be a random number of zeros in regards to the coefficients. So to them they went through and counted all the zeros and there weren't enough. The critics are prescribing how digits should be distributed based on perfect conditions of data collection with no refusals or over sampling. This argument is not realistic.

Were there allegations of falsification of data?

Yes, that was raised as well. The other issue that was raised is that there were too many cases from the South and that was not possible. The samples were also overrepresented White and overeducated respondents. There were also criticisms that "nobody would provide free data to me". That was related to questions about how I could have that data.

How many official inquiries have you had from the Office of Research and what were the approximate timeframes of the inquiries?

There were three inquiries (1999, 2020, and 2021). The individuals kept making allegations and blowing small things into large issues. The allegations were made based on inaccurate information.

In October 2019, you and your co-authors requested withdrawal of five research publications. Why did you make this request?

Once the critics pointed out concerns, I went back into the data and saw that there were issues but not the issues that the critics brought forward. We made corrections to 2 of the 5 papers and the corrections were accepted by the journals. We were in the process of making corrections to the remaining three papers. Before we were able to complete the corrections for remaining papers, these critics raised more concerns. They said that it wasn't possible to observe the results we did with regard to income despite us providing data records showing that our findings were possible and correct. At this point, the critics had posted their objections to various social media boards which created a firestorm of negative reactions (despite) the critics being incorrect in their assessments. After talking with the Dean and my co-authors, we made the decision to withdraw the papers due to the errors in the manuscripts. There were merging errors. One of the issues that arises when using historical data is that some contemporary counties didn't exist in the 1800s and early 1900s. For example, prior to 1824, Leon County was combined with Escambia and Gadsden. So prior to 1824, where would contemporary data for Leon County go? These are the kinds of decisions we had to figure out because historically Leon County did not exist. That is only one example but these were the kinds of decisions we had to address across hundreds of counties. Additionally, there were some errors I made with placing maybe 100 or more individuals in the south instead of the mid-west. Specifically, I inadvertently had Kansas and Missouri coded as being southern. This was one of the issues but we corrected it and the pattern of results didn't change. Another issue was merging some of the data. I merged maybe 8 or so data sources and that's a lot to merge into a single dataset. We found some issues there as well. The merging errors still didn't make any difference in the results. In all the errors we found, the findings didn't change. This was for the lynching papers and as well as the other papers we were making corrections on. Our results were consistent with prior research. All the papers I made corrections on, the general pattern of the findings didn't change.

Why did your co-authors request withdrawal of five of the publications?

FLORIDA STATE UNIVERSITY
OFFICE OF HUMAN RESOURCES

After we talked about, we agreed maybe this was the best route. Initially, we were like maybe we should keep fighting this but ultimately we agreed to withdrawal.

In undertaking the analysis of the data for the five papers, did you have anyone review your analysis or methodologies in analyzing the data?

I shared the analysis with all the co-authors. Sometimes I showed them the output and sometimes I put the results in tables. We talked about different ways to run the statistical analysis. It also depends on what questions we are trying to answer. For each paper, we worked as a team to develop ideas, determine proper methodologies, discuss analyses and findings, and more.

In responding to inquiries regarding these papers, what coding analysis errors did you discover?

Let's say you have a data set and it has individuals and their "attitudes." I'm Interested in examining the way broader context (neighborhoods, schools, and counties) influences those attitudes. In this case, does context influence punitive attitudes? The dataset doesn't necessarily have all of the contextual information added to it. It may have the address or county names for where the sampled individuals live. We have to connect the corresponding contextual information to individuals into a single dataset. We use FIPS codes to connect and merge contextual and individual data into a single file. I merged crime rates, poverty rates, voting data, race and ethnic composition, and other factors to the individual data. The same process was used for lynching. The lynching was a little trickier because if the contemporary county didn't exist you had to use clusters of data. Let's say Leon didn't exist but you had Gadsden county. Where does Leon County values go? These are some of the challenges in trying to link clusters of historical data with current counties and data that can lead to coding errors. This is further complicated because an adequate number of individuals are needed across counties to make meaningful inferences. To conduct this type of analysis and ensure that there are enough people within the county, researchers would create county clusters using U. S. Census County Adjacency File which allows adjacent counties to be collapsed into a larger meaningful county clusters and averages 3-8 counties per county cluster. This is another process I used but I didn't describe the entire process because of limited journal space. Creating county clusters can lead to coding errors because a neighboring county can be merged in multiple clusters.

The Office of Research requested that you develop five chronologies regarding the five withdrawn publications. Did you provide the requested information?

I did provide it.

What about the data from the five publications?

In December 2019, I lost the data and data files for those analyses when my university-issued computer's hard drive crashed. I told our IT person that I was having trouble with my computer. When the IT person attempted to repair my computer, he informed me that the hard drive crashed and he was not able to recover the data from the drive. The college bought me a new hard drive.

In January 2021, a publication (2003) was retracted by the Justice Quarterly (without your request). What was the publication and why did the journal retract the paper?

There was a complaint that there was no way I could have the sample size and the number of schools that I reported and that I coded the urban variable incorrect. The anonymous person raised the complaint to the journal and the FSU Research Office, I responded that it was possible to do what I did.

However, with the attention that it got, the journal's publishers opted to retract the paper, even though I showed that it was possible to obtain the dataset I used in the paper. I was also able to replicate most of the pattern of findings I reported. The journal publishers acknowledged that there was no research misconduct, yet decided to retract the paper. I didn't think it should have been retracted but it was. That data I was using almost 21 years ago came on a CD. The National Educational Longitude Survey. Kids were interviewed from middle school through high school. I used those data when I was a graduate student and into my first job. A number of things have happened since that time. They had to make corrections to the data (not my data set). The number of individuals per school that I used at the time the analyses were conducted aren't acceptable now. Certain methods are no longer used. Part of what I was up against was trying to convey to the journal that the analytical decisions used 20 years ago were standard/acceptable at the time. It wasn't the editors that made the decision, it was the publishers. I was able to replicate what I had in the article fairly closely. Again, not everything was as precise. In my replication, I was able to get a better sample than I did 20 years ago.

Did you provide the journal with the requested information? What happened to the data?

The data were publicly available. I had to rebuild everything. Yet, I provided the journal with the data codes and syntax. I also provided that information to the research office. With regard to my original analyses, I would have purged the original data because it was over 18 years. Now you can get the data off the website.

During a review of your publications by the Office of Research due to allegations of research misconduct, it was report that there were more issues raised about potential errors in the reporting of data in 11 additional papers between (2006-2019). What were the allegations?

There were problems with the standard deviations, they were supposedly off by a couple points for some of the binary variables. In a couple cases, the complainant was wrong which the committee pointed out. In one of the papers, I wasn't responsible for the analysis. In each case, it's the binary variables or making a variable binary. Let's say there is a variable that has five categories (as an example), maybe marital status. Often we might collapse those four or five categories into two categories. In some cases, there may be have missing data and I used a missing data technique that's regression based. It was reported that there were errors in standard deviations, and that was on me. These errors wouldn't impact the findings. The same was noted in the inquiry committee's conclusions by the selected statisticians.

Have the 11 papers with additional concerns been withdrawn?

No, they are still active. The errors that were pointed out are so minute I was actually shocked that this was such a big issue. If that's the case, there would be hundreds of papers retracted. I know with all the controversy that is attached to me that it gets elevated.

Were you able to provide the data for the inquiry committee to review? Why not?

No. The data for the 11 papers were restricted and on my hard drive that crashed. Additionally, some of the publications were more than three years old. In some cases, the papers were 10 years old. I wasn't able to retain data that long.

Is it accurate that in June of 2019 one of your co-authors sent FSU concerns about your publications?

FLORIDA STATE UNIVERSITY
OFFICE OF HUMAN RESOURCES

Yes. That would have been Dr. Justin Pickett. He didn't trust the data and he believed I falsified the data. His role as a coauthor was minimal. He provided voting data and that was it. I brought him on the paper as a favor. He was previously my TA and I was trying to help my student out at the time.

When did you lose your data?

For that paper that Dr. Pickett had concerns about, that paper was published in 2011. I lost data in December 2019.

Did you take steps at that time to ensure your data was backed up?

Yes it was backed up on my laptop. I just kept it on my laptop so it wasn't stored anywhere else. I had never had any issues with storing data on my secured, password protected computed. I'v had files on external drives before and they have failed, I just kept it on my laptop. I felt that once we made corrections everything would be rectified.

Between June 2019 and December 2019, were you aware that you would need to ensure data was available to be reviewed if an investigation was initiated?

Yes, correct. In the inquiry process, the committee did ask for data records. I turned over all of the data records and output files that the Inquiry Committee requested. I was told that the Research Office would maintain the information I provided for seven years. So I thought I was squared away on that front. Also, I still had the data on my computer as I was making corrections.

It has been reported that the Office of the Inspector General at the National Sciences Foundation has initiated an investigation relative to NSF grant #1023337 that was awarded to FSU (2010-2012) and you served as Principal Investigatory? What is the focus of their review and what is the status of that investigation?

I have no idea. They contacted me and what I gather is that they were not happy with FSU's response. They asked FSU for materials and then they came to me to let me know what was going on. This would have been in July of 2021.

The reviewers in the most recent inquiry indicated that the reported errors in analysis were not solely the result of occasional rounding, transcription, coding, or other random errors, but that you were making systematic errors in your analysis. Is this accurate?

That is accurate. While there were possible transcription and other random errors, the misspecification of a dichotomous measure can be a systematic error. One of things I didn't know years ago is that when using regression techniques to deal with missing values, if one doesn't specify whether the measure is dichotomous or not, it can create an error. When I learned the technique a number of years ago, it wasn't an issue. We all learned it the same way. I can assure you that most people don't know about this specification.

Do you have any additional information you would like to add?

No, I would just acknowledge that I made errors that were determined by a number of inquiry groups to be insignificant to the findings of the research and, most importantly, not research misconduct. I should have been more detailed in how I presented information. This has been a tough and trying process for me and my family.

FLORIDA STATE UNIVERSITY
OFFICE OF HUMAN RESOURCES

_____

By signing this document I am confirming that I have been given the opportunity to review these notes; that the contents reflect, to the best of my knowledge, an accurate description of the incident(s) or subjects described; and that I have given these statements in good faith.

6/14/2022 | 8:36 PM EDT

Acknowledged by: ___*Eric Stewart*___ Date: ___/___/_____
D99B455A1421491...
Eric Stewart

6/14/2022 | 2:40 PM EDT

Notes Prepared by: ___*Rebecca Peterson*___ Date: ___/___/_____
C3BEFEB4759A481...
Rebecca Peterson

Peer Panel Document
Re: Eric Stewart

#13

Overview
There has been a growing interest in developing techniques and applying different strategies to identify problematic research. These approaches are typically quite straightforward, can be calculated/applied relatively easily, and are excellent at differentiating problematic research from nonproblematic research. Below is a description of some of these techniques and how they are useful in identifying erroneous findings. Each of these indicators in isolation is accurate at detecting erroneous research, but when more than one is used in combination, then the overall accuracy of detecting erroneous research is increased. To err on the side of caution, the following analysis focuses on multiple indicators of problematic research and then applies these indicators across a number of Dr. Stewart's articles to determine whether the studies appear to be incorrect and/or problematic. In situations where the violation was likely to have occurred, but could not be determined with complete confidence, it was not coded as a violation. Last, not all of the indicators could be assessed in each of the articles. For example, those publications where descriptive statistics were not provided, it was not possible to check the standard deviations for accuracy.

Benford's Law
Benford's law captures the logarithmic distribution of the first digit, but an extension of this law has been used to help identify results that are likely due to incompetence. Specifically, the terminal digit (i.e., the last value reported in a decimal) has been shown to be uniformly distributed. What this means is that there should be relatively equal numbers of all values (i.e., zero through nine, each occurring about 10% of the time), but research that has been shown to be incorrect has deviated from this pattern. Specifically, there are certain values (e.g., zero) that are not used as frequently by incompetent researchers. So, it is relatively easy to examine the terminal digit and see whether the values deviate from a uniform distribution. If they do, then this would be a red flag of potentially problematic research. This approach has been used widely to identify erroneous data as it is straightforward, easy to calculate, and one of the most accurate tests for problematic research.

Incorrect Standard Deviations
Another approach to identify problematic research is through the estimation of standard deviations for binary variables. The equation for the standard deviation for a binary variable is straightforward. Specifically, the equation for the standard deviation with binary variables is as follows:
$$SD = \sqrt{(N/(N-1) \times P(1-P)}$$
This equation can easily be estimated as long as the sample size (N) and the mean/proportion of the sample coded as "1" is provided. Problematic research is frequently detected by applying this equation as incompetent researchers frequently report standard deviations that do not conform to this easy-to-estimate equation.

Stability in Coefficients and Standard Errors
In social science research (including criminological research), the most widely used approach to analyzing data is some type of multivariate statistical modeling. Put simply, multivariate modeling allows for the estimation of an association between two variables (e.g., X and Y) while holding constant the effects of other variables (e.g., W). This approach is useful as anytime there is an association between X and Y, then it could be explained by other variables (e.g., W) that covary with them. By including controls for these other variables (e.g., W), then it allows for a more accurate estimation of the true nature of the association between X and Y. What is important to realize is that if the association between X and Y is first estimated and then in a subsequent model additional variables (e.g., W) are introduced and those additional variables overlap with X and Y, then they will change the association

between X and Y. This change could be small and relatively insignificant, or it could be large and produce very different results.

For example, suppose the bivariate association between X and Y is found to be r = .20. Then in a subsequent (multivariate) model another variable, W, is introduced into the statistical equation. If W is related to X and Y (and this would be the *a priori* reason for including W; if W is unrelated to X and Y, then there is no mathematical reason to include it in the statistical model and it can be eliminated from the equation), then the correlation between X and Y will likely change. In this case, perhaps it will drop to r = .18. Typically, as more variables are entered into the equation, the greater the probability that it will affect the association between X and Y.

The association between two variables is partially a function of the coefficient (i.e., the value reported; in this example, the coefficient would be "r" or the correlation) and what is known as the standard error (the standard error was not provided in this example). Tests of statistical significance (for the coefficient measuring the association) are determined by dividing the coefficient by the standard error. So, the larger the coefficient in relation to the standard error, the greater the probability of detecting a statistically significant association. The key point to bear in mind, though, is that coefficients and standard errors typically change across models as additional variables are introduced into the models.

An additional point is important to make. Oftentimes researchers are interested in examining what is called an interaction. Interactions (from a statistical approach) are created by examining whether the effects of two variables (say, X and Q) predict an outcome (say, Y). Usually, to test for interactions, an additive model is first calculated that simply examines whether X and Q would have independent additive effects on Y. Then, the interaction term is introduced (which is created by multiplying X and Q to create a new term, XQ). The interaction model thus includes X, Q, and XQ. Since XQ is a multiplicative function of X and Q, the coefficients and standard errors generated in the additive model for X and Q typically expand greatly in the interaction model. Indeed, there are debates on how best to deal with the "explosion" of standard errors that often results when comparing an additive model to an interactive model. (This is the result of what is known as collinearity or multicollinearity and is often viewed as a problem in multivariate research and should always be evaluated prior to estimating multivariate models. In interactive models, though, collinearity is a known problem because XQ is highly correlated with and a mathematical function of X and Q. As a result, the best way to deal with collinearity in interactive models is debatable [e.g., through mean centering, doing nothing, etc.]. What is important to underscore, however, is that the coefficients and standard errors in interactive models become more unstable and thus coefficients and standard errors are expected to change between the additive model and the interactive model. If they do not, then this would be a red flag for potentially problematic research and erroneous results.

In situations where the results of a series of multivariate models are presented, it is possible to examine whether the coefficients and standard errors fluctuate across the models. For the most part, it is typical to see coefficients and standard errors change across models, sometimes drastically and other times relatively little. However, it would be a statistical anomaly to see complete stability in coefficients and standard errors across multiple models where additional variables are introduced into those equations. And, it would be impossible to achieve stability in coefficients and standard errors when comparing additive models to interactive models. If the coefficients and standard errors are found to be extremely stable across models, then this would be a cause for concern and a potential indicator of an incompetent researcher estimating the statistical models.

Other Indicators, Errors, and Irregularities
There are a host of other errors and irregularities that have been shown to cut across Dr. Stewart's publications. For example, p-values associated with tests of statistical significance for coefficients should be uniformly distributed, but in some of his studies they deviate drastically from this distribution. Moreover, in several of his papers there are errors in bivariate hypothesis tests, problems with the sample sizes, and issues with descriptive data. These errors and irregularities cannot be the result of coding and/or rounding errors as they are impermissible solutions when compared to other values reported within his studies.

Summary
A single study having even one of these indicators of incompetence would be a significant cause for concern that warrants additional scrutiny. Indeed, a thorough search of the literature has failed to uncover any instances of where a study having one of these indicators was later cleared of any problems. Having multiple problem indicators would be about as clear of an indication as possible that the study under question contains significant errors and impossible solutions, indicating the work of an incompetent researcher.

As Table 1 makes clear, sixteen (16) of Dr. Stewart's studies were found to have problematic indicators. These problematic indicators were found across numerous articles and span at least 17 years of his academic career, including the entire time he has been employed as a faculty member in the College of Criminology and Criminal Justice at Florida State University. These were no errors that could be explained away by human error or analytical decisions as they were detected across multiple studies using different data and that employed a variety of methodological and statistical approaches. In short, the problematic indicators were found to be pervasive in his studies, were not confined to just a single problematic indicator, and were detected across a lengthy stretch of his academic career providing compelling documentation of flagrant research incompetence.

Table 1. Summary Table for Indicators of Impossible Results and Erroneous Statistics across Multiple Stewart Articles

Year	Benford's Law	Standard Deviation	Stability in Coef./SEs	Other Indicators
2003[i]	X			X
2004[ii]				X
2006[iii]	X	X		
2006[iv]	X			
2009[v]		X		
2010[vi]		X	X	
2011[vii]	X	X	X	X
2012[viii]		X		
2013[ix]	X	X		
2015[x]	X	X	X	X
2017[xi]		X		
2017[xii]	X	X		
2018[xiii]	X	X	X	X
2019[xiv]	X	X[a]	X	X
2019[xv]	X	X	X	X
2020[xvi]		X		

Notes:

SE = standard error;

X = violation observed

[a] = corrected this error after initial online publication

[i] Stewart, E. A. (2003). School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly, 20*, 575-604.

[ii] Stewart, E. A., Elifson, K. W., & Sterk, C. E. (2004). Integrating the general theory of crime into an explanation of violent victimization among female offenders. *Justice Quarterly, 21*, 159-181.

[iii] Stewart, E. A., Schreck, C. J., & Simons, R. L. (2006). 'I ain't gonna let no one disrespect me': Does the code of the street reduce or increase violent victimization among African American adolescents? *Journal of Research in Crime and Delinquency, 43*, 427-458.

[iv] Stewart, E. A., & Simons, R. L. (2006). Structure and culture in African American adolescent violence: A partial test of the "code of the street" thesis. *Justice Quarterly, 23*, 1-33.

[v] Stewart, E. A., Baumer, E. P., Brunson, R. K., & Simons, R. L. (2009). Neighborhood racial context and perceptions of police-based racial discrimination among black youth. *Criminology, 47*, 847-887.

[vi] Stewart, E. A., & Simons, R. L. (2010). Race, code of the street, and violent delinquency: A multilevel investigation of neighborhood street culture and individual norms of violence. *Criminology, 48*, 569-605.

[vii] Johnson, B. D., Stewart, E. A., Pickett, J., & Gertz, M. (2011). Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. *Criminology, 49*, 401-441.

[viii] Warren, P. Y., Stewart, E. A., Tomaskovic-Devey, D., & Gertz, M. (2012). White's residential preferences: Reassing the relevance of criminal and economic stereotypes. *Race and Justice, 2*, 231-249.

[ix] Mears, D. P., Stewart, E. A., Siennick, S. E., & Simons, R. L. (2013). The code of the street and inmate violence: Investigating the salience of imported belief systems. *Criminology, 51*, 695-728.

[x] Stewart, E. A., Martinez, R., Baumer, E. P., & Gertz, M. (2015). The social context of latino threat and punitive latino sentiment. *Social Problems, 62*, 68-92.

[xi] Hughes, C., Warren, P. Y., Stewart, E. A., Tomaskovic-Devey, D., & Mears, D. P. (2017). Racial threat, intergroup contact, and school punishment. *Journal of Research in Crime and Delinquency, 54*, 583-616.

[xii] Mears, D. P., Stewart, E. A., Warren, P. Y., & Simons, R. L. (2017). Culture and formal social control: The effect of the code of the street on police and court decision-making. *Justice Quarterly, 34*, 217-247.

[xiii] Stewart, E. A., Mears, D. P., Warren, P. Y., Baumer, E. P., & Arnio, A. N. (2018). Lynchings, racial threat, and whites' punitive views towards blacks. *Criminology, 56*, 455-480.

[xiv] Mears, D. P., Stewart, E. A., Warren, P. Y., Craig, M. O., & Arnio, A. N. (2019). A legacy of lynchings: Perceived black criminal threat among whites. *Law & Society Review, 53*, 487-517.

[xv] Stewart, E. A., Johnson, B. D., Warren, P. Y., Rosario, J. L., & Hughes, C. (2019). The social context of criminal threat, victim race, and punitive black and latino sentiment. *Social Problems, 66*, 194-221.

[xvi] Warren, P. Y., Cochran, J., Shields, R. T., Feldmeyer, B., Bailey, C., & Stewart, E. A. (2020). Sentencing departures and female defendants: Assessing the effects of racial and ethnic threat. *Crime & Delinquency, 66*, 59-92.

Peer Panel Document

Re: Eric Stewart

#14

Eric Stewart Interview Review

Overview

Identifying problematic research can be difficult, but relatively simple techniques have been developed that can be used to help isolate studies that should be "red flagged" and that warrant additional attention. The key point to remember—and one that is often lost—is that these techniques focus on within-study anomalies or peculiarities. That is, there are certain patterns in results that should be consistent within each study. If those patterns are not observed, then systematic errors are a likely explanation. This is different than between-study results wherein the results of one study may be quite different from the results of another study. For the most part, between-study differences provide virtually no information about errors as those differences can be due to any number of legitimate reasons. So, the focus is largely on within-study patterns when examining studies for systemic errors.

In addition, another issue with trying to identify problematic research is that the data are not available for examination (either because the data cease to exist or because researchers do not want to provide the data employed to create their questionable results). Without having access to the data, it would seem quite difficult to determine research that is valid versus research that is not. Again, however, researchers have identified relatively simple tactics that focus on within-study patterns of results that assist in identifying potentially problematic research. These simple tactics do not require having the actual data and thus are used widely when trying to determine whether a study is indeed questionable.

One of the most common ways to identify problematic research is through the estimation of standard deviations for binary variables. The equation for the standard deviation for a binary variable is straightforward and can be calculated when other descriptive information is provided. Fraud-detection researchers have used this simple test widely as research that has been shown to be questionable consistently reports standard deviations for binary variables that are incorrect. The only reasons standard deviations would be incorrect is through 1) a human error in reporting the standard deviation (e.g., incorrectly copying the value from output to table) or 2) creating false values intentionally. When multiple standard deviations are reported incorrectly, then it is either because of research incompetence or intentional misrepresentation.

The second straightforward approach to identifying potentially problematic research is by applying what is known as Benford's Law. Put simply, Benford's Law states that 2nd and 3rd digits beyond a decimal should be uniformly distributed. What that means is that there should be approximately the same number of zeroes, ones, twos, threes, fours, fives, sixes, sevens, eights, and nines of the last reported digit in a decimal. Researchers with problematic data and results typically stay away from certain values when reporting results (especially zeroes). This has been shown in research time and again. As a result, by simply tallying the values in the last digit place reported (usually the 2nd or 3rd digit is reported), it is easy to determine whether the pattern of values corresponds to what would be expected (relative uniformity in values). Deviations from a uniform distribution have been shown to be quite a strong indicator of problematic research.

Prepared by: Dean Thomas Blomberg   Date: 10/6/2022

These detection techniques are by no means the only ones available, but they are two of the most widely used and easiest to calculate. Studies that are flagged after using these techniques are typically viewed as warranting additional attention and scrutiny. Sometimes additional errors and anomalies are uncovered and oftentimes the data (when available) are analyzed to determine whether the results can or cannot be replicated.

Point-by-Point Responses (talking point) to Dr. Stewart's Answers to the Interview Questions.

What is your training in statistical analysis?

Answer: Pretty extensive.

Talking point: What is meant by pretty extensive. This is a vague response that tells nothing about Dr. Stewart's training in relation to the statistics that he used and his ability to manage data and create large-scale datasets. If his training in statistical analysis was extensive, then there is no way that he would have created the errors that he did in his studies. Calculating the standard deviation, for example, is something that is taught in middle and high schools and yet he repeatedly miscalculated and/or misreported the standard deviation. If he is unable to estimate the standard deviation correctly, then serious concerns should be raised about his competence to estimate other, more advanced statistical techniques.

Are you familiar with FSU Policy 7A-26?

Answer:…In some data that we use, we don't keep it longer than two years because of the sensitive nature.

Talking point: Data are typically retained forever and whether data are considered sensitive is irrelevant. All data are considered sensitive and data that contain extremely sensitive data do not have requirements on length of keeping the data, but rather with how the data are housed and stored. Typically, sensitive data are password-protected, housed on computers without a connection to the internet, and use programs and other techniques to avoid the data being accessed by unauthorized users. For someone to say that data are only retained for a couple of years is completely misleading. It can take years for data to be collected and it can take decades for the data to be thoroughly analyzed and studies published from it. It would be difficult to find any criminological sample that was completely erased after any amount of time let alone two years. Dr. Stewart should be well aware of these practices as he has worked extensively with the FACHS data, data that were collected in 1996 and that he still analyzes in his own research. Moreover, during his time as a graduate student, the FACHS data were being collected by his major professor (Ronald Simons) and so he would have had experience with handling sensitive data and how to retain it.

It was reported that there were concerns regarding several of your publications which prompted an inquiry from the Office of Research.

Answer: His response focused on zeros in the coefficients which he claimed he did not understand.

Talking point: This is Benford's Law. Dr. Stewart's research studies were found not to conform to Benford's Law which is one of the most salient red flags for potential problems. Keep in mind, that it was not just one of his studies that did not conform to Benford's Law, but almost all of them that are in question. Moreover, he seems to downplay this concern by saying that there were not enough zeroes and that he took the coefficients out three decimal places. First, incompetent researchers typically shy away from including zeroes so when Benford's Law is violated and the violation has to do with a deviation for zeroes (rather than any other number), then this is even a bigger signal that something is wrong. Second, carrying out the values three decimal points has nothing to do with the pattern of results as Benford's Law is more applicable to the 2nd and 3rd digit than the first. And, last, Dr. Stewart is incorrect in noting that there should be a random number of zeroes in regards to the coefficients. There should not be a *random* number of zeroes, but rather a *uniform* number of zeroes according to Benford's Law. His research did not have a uniform number of zeroes indicating a deviation from Benford's Law. His responses demonstrate his research incompetence.

Were there allegations of falsification of data?

Answer: Yes, that was raised as well. There were too many cases from the South and that was not possible and that "nobody would provide free data."

Talking point: This is not entirely true or at least taken out of context. Once red flags were raised about potential problems due to standard deviations being incorrect and deviations from Benford's Law, a more critical analysis was undertaken. During this investigation additional areas of concern were raised. One of these concerns was with the sample sizes across the different geographical errors. Another was that collecting large datasets is expensive and there has to be some type of funding source in order for the data to be collected. So, it was not that nobody would provide free data (free data are provided all of the time; search the ICPSR website), but rather that collecting data is expensive and so without a funding source data would likely not be collected on a large-scale (or even small-scale) basis. Not having a funding source for these data raises additional concerns over the existence of the data and how they were collected without funding.

In October 2019, you and your coauthors requested withdrawal of five research publications.

Answer: Merging errors.

Talking point: Merging errors do not create within-study anomalies that are used to detect problems. Merging errors might lead to incorrect findings and be indicative of between-study

Prepared by: Dean Thomas Blomberg   Date: 10/6/2022

differences. Also, Dr. Stewart's explanation suggests that this was not a merging error, but rather an error in his decision on how to code/classify south versus Midwest. According to Dr. Stewart, once he corrected his classification error, the results did not change. It would be interesting to see the actual data as multilevel models (which he used) are not always stable when moving people from one region to another. Attributing his classification error as a merging error is really deceptive for those who are trained in statistics. Merging errors can occur because of a software malfunction that is no fault of the researcher (though it is the researcher's job to check to make sure that a merging error did not occur). In this case, the error that Dr. Stewart created was of his own doing and through classification decisions that he made again documenting incompetence..

Why did your coauthors request withdrawal of five of the publications?

Answer: We agreed this was the best route.

Talking point: Any researcher in this situation would turn over their data and allow others to see what they did, how they did it, and whether what they did could be replicated. All that needed to be done was to release the data. Agreeing to five retractions when Dr. Stewart stands by his findings and data is beyond comprehension for any researcher.

Did you have anyone review your analysis or methodologies in analyzing the data?

Answer: I shared the analysis with all the coauthors by showing the output and tables.

Talking point: Showing output and tables is not showing the data. Output can be edited and the output provides the tables. It would be considered extremely odd that none of his coauthors ever saw the data and that they only were provided with output and tables. Again, no responsible and competent researcher would withhold data from co-authors.

In responding to inquiries regarding these papers, what coding analysis errors did you discover?

Answer: He does not answer the question.

Talking point: He did not answer the question. So, he allowed five papers to be retracted even though he did not identify any coding errors? This makes no sense. He makes it seem as though he did not fully describe his coding process which is true, but that is not an error and that is not a reason for a retraction. Again, another example of incompetence.

What about the data from the five publications?

Answer: I lost the data file for those analyses that was destroyed on my laptop.

Talking point: Deleted files can be recovered and there are programs available that allow for files to be recovered from hard drives that crash. Forensic approaches are extremely effective in recovering files. A data file that is "lost" while it is under investigation is particularly peculiar. Three outcomes are potentially possible. First, the data could have been lost (least likely explanation). Second, the data may have never existed. Third, the data existed in some format but were destroyed to prevent others from accessing them. This latter possibility is bolstered by the fact that Dr. Stewart asked a colleague about a data-eraser program that deletes temporary data/files/backups from computers that analyze restricted-use samples. This request for a data-eraser program can be validated by the faculty colleague who was asked about data-eraser programs.

In January 2021, a publication was retracted by the *Justice Quarterly*. What was this publication and why did the journal retract it?

Answer: Dr. Stewart noted that the sample sizes and number of schools was incorrect and that certain methods were no longer used.

Talking point: The issue of sample size is straightforward and is easily checked. It was checked by *Justice Quarterly* and did not match Dr. Stewarts' sample size. Dr. Stewart mentions that certain methods are no longer used. What are those methods? Methods are not typically discarded and not used anymore. Certainly newer approaches might be introduced but that does not make the previous methods obsolete. Being vague and imprecise allows him to create a moving target where he cannot be pinned down on any specifics. He also notes that he was able to replicate what he reported and that he was able to get a better sample. If he replicated the results, then it would be odd that the journal would retract it. In addition, how would he be able to get a "better sample." These data are secondary and thus the sample would be provided to him and he would have no bearing on the quality of the sample. His response is completely misleading and again shows research incompetence.

Did you provide the journal with the requested information? What happened to the data?

Answer: I would have gotten rid of the original data because it was so many years ago.

Talking point: Deleting data where it is not required is unheard of and not a best practice. There is always the possibility that questions might be raised about a study or that another study will be published off the original data. This answer is at best alarming and another example of incompetence.

Were you able to provide the data for the inquiry committee to review? Why not?

Answer: Some of the publications were more than two or three years old. I wasn't able to retain data that long.

Talking point: This makes no sense. Data are usually retained for decades (or longer). After collecting data, a researcher typically publishes from that sample and can analyze the data as part of their career for decades. Two or three years in terms of publishing is a blink of an eye and really unheard of – once again demonstrating incompetence.

When did you lose your data?

Answer: For that paper that Mr. Pickett had concerns about, that paper was published in 2011. I lost the data in December 2019.

Talking point: Dr. Stewart continuously states that he deletes data after 2-3 years. Yet, here, in his own words, he retained the data for at least eight years. He only lost the data after he was being investigated and after he was being asked to provide the data. Very troubling at the least.

Did you take steps at that time to ensure that your data was backed up?

Answer: Yes, it was backed up on my laptop. I've had external drives before and they have failed.

Talking point: Keeping data on a laptop is not backing it up. Backing up data means that it is kept in multiple places. Dr. Stewart had previous problems with failing drives and so he should have been acutely aware of the importance of keeping the data in multiple locations. It is a best practice to have the data stored on a work computer, a cloud drive, and an external hard drive when legally permitted. There does not appear to be any legal impediments to Dr. Stewart keeping his data in multiple places. Having data in only one place is absurd and virtually unheard of today. Again, this failure to adequately backup his data further documents incompetence.

Between June 2019 and December 2019, were you aware that you would need to ensure data was available?

Answer: Yes, I turned over materials and records to them.

Talking point: Dr. Stewart did not turn over the raw data. He turned over some output and some other information, but not the data. Had he turned over the data, then all of this would have been resolved quickly.

The reviewers in the most recent inquiry indicated that that the reported errors in analysis were not solely the result of occasional rounding… Is this accurate?

Answer: That is accurate. One of the things I didn't know when using this technique is you have to specify whether the variable is dichotomous or not. If you don't, then it can create an error. I can assure you that most people don't know that.

Prepared by: Dean Thomas Blomberg   Date: 10/6/2022

Talking point: This is only partially true. With multiple imputation (MI) which it seems that he is talking about, dichotomous variables must be specified. However, this is actually built into the programs and it makes you specify whether the variable is dichotomous. The equations for MI make this clear. Even if (for some reason) the researcher did not specify that they were working with a dichotomous variable, it is unlikely that it would produce significant errors in the values. Plus, it should still create within-study consistencies that would not be red-flagged when fraud-detection techniques were applied to the study. Last, anyone using MI knows (or should know) that dichotomous variables have to be specified. This would be taught in any intro class, book, or article dealing with MI and it would be inherently built into the software used to calculate it. Again, documentation of incompetence.

Conclusions

At every step in the research inquiry and investigation process, Dr. Stewart's studies (in question) have been red-flagged. Upon closer inspection, they have been found to have even more signs of errors and misrepresentations. When pressed to provide the data, Dr. Stewart failed to give the raw data to his coauthors. In short, nobody has seen the raw data. Before FSU was able to access the data, Dr. Stewart lost the data on his laptop and any backups of it. According to Dr. Stewart, it is his policy to retain data for 2-3 years (though his own actions contradict this practice). What is interesting is that all researchers—especially those who have pretty extensive training in statistics—use (or should use) what is known as syntax for their data creation, data management, and statistical analysis. Syntax is a written language of commands which provides a written step-by-step notation of all decisions made during the research process. This allows for a quick way to remember what was done and it also allows for the analyses to be reestimated with a click of the keyboard. Moreover, syntax is not considered sensitive so it can be publicly disseminated without violating any data-protection rules/laws governing sensitive data. Even if Dr. Stewart had lost his data, he should have had his syntax available that would have allowed him to recreate his dataset and analyses within a relatively short period of time.

Peer Panel Document
Re: Eric Stewart

#15

THE FLORIDA STATE UNIVERSITY
OFFICE OF THE PROVOST

March 10, 2023

Dr. Eric Stewart
College of Criminology and Criminal Justice
Florida State University

Dear Dr. Stewart:

This letter serves to officially notify you that Florida State University (FSU) intends to terminate you from your position of Professor on or after March 30, 2023 for incompetence and negligence in the performance of your duties. This action is based on the facts outlined below:

Beginning in May 2019, concerns were raised regarding data irregularities for several published papers for which you were either the lead author or co-author. When concerns about the papers were originally raised, they were accompanied with allegations of data manipulation, data fabrication and/or fraud. In October 2019, you and your co-authors voluntarily requested withdrawal of five research publications from the journals of *Criminology*, *Social Problems*, and *Law and Society Review*. These articles included:

1. Johnson, B. D., Stewart, E. A., Pickett, J., & Gertz, M. (2011). Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. *Criminology*, 49, 401-441.
2. Stewart, E. A., Mears, D. P., Warren, P. Y., Baumer, E. P., & Arnio, A. N. (2018). Lynchings, racial threat, and whites' punitive views towards blacks. *Criminology*, 56, 455-480.
3. Stewart, E. A., Martinez, R., Baumer, E. P., & Gertz, M. (2015). The social context of latino threat and punitive latino sentiment. *Social Problems*, 62, 68-92.
4. Stewart, E. A., Johnson, B. D., Warren, P. Y., Rosario, J. L., & Hughes, C. (2019). The social context of criminal threat, victim race, and punitive black and latino sentiment. *Social Problems*, 66, 194-221.
5. Mears, D. P., Stewart, E. A., Warren, P. Y., Craig, M. O., & Arnio, A. N. (2019). A legacy of lynchings: Perceived black criminal threat among whites. *Law & Society Review*, 53, 487-517.

Your communications with the three journals regarding the withdrawals indicated that the decision to withdraw the five publications came from all the co-authors. You further indicated that because you were responsible for the data and analyses for the five publications you took the lead in responding to the allegations of data manipulation, data fabrication, and/or fraud. You ultimately drafted and signed the requests for withdrawals of the five papers: On October 24, 2019, you sent a request to *Criminology* retracting two articles and a request to *Law and Society Review* retracting one article and on October 25, 2019, you sent a request to *Social Problems* retracting two articles. In three withdrawal letters,

while denying any fraud, you admitted that numerous analyses errors had occurred, warranting withdrawal of the papers; specifically, you stated:

"*Contrary to assertions in social media and other online outlets, there was no falsification, fabrication, or other fraud associated with these papers or others that relied on the public opinion and contextual data. However, questions have been raised about the data and results reported in these papers. I have taken the lead in responding to these questions because I undertook the analyses, but this request for withdrawal comes from all of the co-authors. In the course of responding to the questions, it became clear that the analyses errors that included coding mistakes and transcription errors, exceed what the co-authors and I ultimately concluded as acceptable for a published paper. For this reason, on behalf of the co-authors, I request that the papers be withdrawn.*"

On January 21, 2021, the Editors and Publisher of *Justice Quarterly* retracted your sole-authored 2003 paper, published while you were at Georgia State University (GSU):

> Stewart, Eric A. 2003. School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly* 20:575-604.

The journal's published retraction commentary explained the unilateral retraction as being based on errors in the article and the inability of referees to replicate the results, casting doubt over "the reliability of the results and conclusion" and specifically read, in relevant part: "*In response to concerns raised about this article, the Editors and the Publisher provided the author with the opportunity to respond to the questions raised about this article, and commissioned a review by three independent referees of the original article and the author's response to the concerns.*

The independent review concluded that some of the concerns raised about the article could possibly be attributed to analytic and coding decisions, but due to possible errors in select analyses, as well as the inability of the referees to replicate the article's exact sample and models, there was doubt cast over the reliability of the results and conclusions in this study. For this reason, we have made the decision to retract the article. The author has been informed of this decision."

In addition to these six retracted articles, concerns were also raised regarding potential reporting of data in 11 additional papers authored by you. Per FSU Policy (7A-2) regarding misconduct in research, FSU charged an Inquiry Committee to conduct an initial review regarding these additional 11 papers to determine if a full Investigation was warranted. You informed the Inquiry Committee that the data for these papers had been either lost in a hard disk crash or that the data were no longer available because you routinely delete your data within two to three years because of its sensitivity. However, in the interview you also stated that on occasion you had kept some data nine years, which seems to contradict the other information you provided to the Inquiry Committee. Because of your negligence in backing up your data, that data were no longer available and, therefore, FSU was unable to provide a comprehensive review of the allegations regarding the 11 additional articles in order to determine if there were falsifications, misrepresentations, and/or fraud (see References for full citation of these

articles). Further, there was consensus from the members of the Inquiry Committee that someone sophisticated enough to routinely use a desktop computer, laptop computer and iPad understands that computer systems can periodically fail and, thus, that safeguards and backups must be implemented. Given the accusations you faced, failing to back up the data that was being questioned as part of a Research Misconduct Inquiry was extremely careless, and demonstrates gross negligence and incompetence.

To date, three separate Research Misconduct inquiries (2019, 2020, 2021) have been conducted by FSU over various allegations of fraud and falsification involving: (1) the five withdrawn publications, (2) the retracted 2003 publication that you authored while a faculty member at GSU, and (3) the additional 11 papers published between 2006 and 2020. GSU also conducted an inquiry relative to the 2003 publication. The FSU and GSU inquiries were limited in their effectiveness because you again indicated that the relevant data to evaluate the veracity of your data analyses were no longer available. As you know or should know, the undeniable and fundamental obligation of all researchers is to maintain data to allow third part replication. However, as you indicated repeatedly to questions about access to your data, you either (1) deleted data within two to three years because of its purported sensitivity, or (2) you lost the data because of a computer crash and your failure to have appropriate backup mechanisms in place to safeguard the data. Providing output and tables without the actual data from which the output and tables were derived prevents third party replication and is directly contrary to accepted scientific research standards (i.e., National Institutes of Health, Conduct of Research 7th Edition, 2021:10).

Regarding the six retracted papers (five withdrawn voluntarily by you and your co-authors and one by the publisher), to date the FSU Research Misconduct inquiries (2019, 2020 and 2021) have not led to a satisfactory resolution, as noted above. You have not been exonerated of wrongdoing, found to have engaged in misconduct, or recommended for a full investigation. The Inquiry Committees, composed of faculty members charged to look at specific instances of your retracted papers and any related misconduct, have been unable to reach a definitive outcome because they have not obtained from you (or elsewhere) the necessary data to make a final conclusion as to whether research misconduct has occurred. This uncertainty has placed you, the College of Criminology and Criminal Justice, its faculty, its students, FSU, and the broader Criminology research community in a problematic and uncertain position.

To reach closure, and to further investigate possible policy violations outside of the scope of the University's Research Misconduct policy, the Office of Human Resources reviewed the information and investigated whether or not you violated any other policies at the University. As part of this investigation, on April 11, 2022, you met with Rebecca Peterson, Sr Associate Director, Faculty Relations and Dean Thomas Blomberg, College of Criminology and Criminal Justice, for an investigatory interview. Following that interview, a review of your responses was conducted. In the review, it was found that your responses were often vague, inaccurate, misleading, and/or noncomprehensible. For example, when you were asked about your training in statistical analysis, you responded that it was pretty extensive. However, if that was true, then it seems implausible that you

would have created the mistakes that you did in your studies, such as the repeated errors in merging and coding data. Further, in attempting to explain your failure to properly maintain data, you noted that you had stored the data on a laptop. However, backing up data on a laptop is not backing up data. Standard research practice dictates that backing up data would include keeping data in multiple places such as a work computer, a cloud drive, and an external hard drive, if permitted. Also, you indicated that you often delete your data after two to three years because of sensitivity. Numerous researchers employ sensitive data that they do not delete but rather maintain following established or required protocol to protect the data. Moreover, most researchers keep data for decades in order to reanalyze or share the data. Further, you indicated that you provided the Inquiry Committee with information (namely output and tables), but you failed to provide the Inquiry Committee with the data for the output and tables. Therefore, the three Misconduct Committees could not reach a conclusion about the research misconduct allegations that were lodged against you as you failed to provide access to the data that the questioned studies were based upon.

Based on the totality of the information and the glaring demonstration of data mismanagement and the unprecedented number of articles retracted, I find that the evidence supports a termination for incompetence and negligence in the performance of your duties. As a published researcher and a tenured full professor, you are responsible for exercising competence in basic research processes, and competence in data management, and competence in the basic parameters of the statistical analyses you employed. <u>You are fully responsible for the integrity of the results that are generated</u>. Your conduct has adversely impacted and reflected poorly on the University, the College of Criminology and Criminal Justice, its faculty and students, and the discipline of Criminology. This includes numerous expressed reservations from persons who were considering joining the College, i.e., faculty and PhD prospects, questioning of our PhD students who are on the job market about the status of Stewart's research misconduct allegations and about the integrity of their own data, numerous inquiries and stated concerns from our criminological peers around the country, potential negative consequences to the careers of your co-authors and students, questioning from current students about your credentials to teach given the misconduct allegations against you, and an overall cloud over FSU and the College in relation to the negative social media responses to the unresolved allegations of research misconduct against you. As Provost, I must uphold the values and the mission of the university and, therefore, I have decided to proceed with the notice of termination.

Please be advised that within twenty (20) days of receipt of this notice, you may submit a response in writing for consideration before a final action is taken. You may also include in your response supporting materials from other individuals. You also have the right to request a meeting with me to refute the statements contained herein.

As an alternative, you may request a review by an appropriate faculty committee (hereinafter referred to as the "Peer Panel") prior to the issuance of the Notice of Termination. This Peer Panel would serve as the peer review panel specified in Article VI, Section B(8)(c)-(f) of the Florida State University Constitution and the Regulations of the Florida State University, FSU-1.004 (IV)(B)(8)(c)-(f). Any request for a meeting or a Peer Panel must be made, in writing, within ten (10) days of receipt of this

Notice of Intent to Terminate, to Rebecca Peterson, Faculty Relations, Human Resources, 6110 University Center, Florida State University, Tallahassee, Florida 32306. Ms. Peterson can also be reached at (850) 645-2202.

As an official representative of Florida State University, I assure you that it is our sincere desire to take only warranted disciplinary action against you, based on true and correct charges. Therefore, we are sincerely interested in receiving and considering your response before taking the disciplinary action to terminate you.

Sincerely,

James J. Clark, Ph.D.
Provost and Executive Vice President
Florida State University

I understand and acknowledge receipt of this notification.

_____   _____
Eric Stewart   Date

Cc: Employee Personnel File (Stewart, E.)
        Thomas Blomberg, Dean, College of Criminology
        Faculty Development and Advancement
        Faculty Relations

References

Berg, M.T., Stewart, E.A., Brunson, R.K., & Simons, R.L. (2012). Neighborhood cultural heterogeneity and adolescent violence. *Journal of Quantitative Criminology, 28,* 411-435.

Berg, M.T., Stewart, E.A., Intravia, J., Warren, P.Y., & Simons, R.L. (2016). Cynical streets: Neighborhood social processes and perceptions of criminal *injustice. Criminology, 54*(3), 520-547.

Berg, M.T., Stewart E.A., Schreck, C.J., & Simons, R.L. (2012). The victim-offender overlap in context: Examining the role of neighborhood street culture. *Criminology, 50*(2), 359-390.

Berg, M.T., Stewart, E.A., Stewart, E., & Simons, R.L. (2013). A multilevel examination of neighborhood social processes and college enrollment. *Social Problems, 60*(4), 513-534.

Mears, D.P., Stewart, E.A., Siennick, S.E., & Simons, R.L. (2013). The code of the street and inmate violence: Investigating the salience of imported belief systems. *Criminology, 51*(3), 695-728.

Stewart, E.A., Baumer, E.P., Brunson, R.K., & Simons, R.L. (2009). Neighborhood racial context and perceptions of police-based racial discrimination among black youth. *Criminology, 47*(3), 847-887.

Stewart, E.A., Schreck, C.J., & Simons, R.L. (2006). "I ain't gonna let no one disrespect me": Does the code of the street reduce or increase violent victimization among African American adolescents? *Journal of Research in Crime and Delinquency, 43*(4), 427-458.

Stewart, E.A., & Simons, R.L. (2006). Structure and culture in African American adolescent violence: A partial test of the "code of the street" thesis. Justice Quarterly, 23(1), 1-33.

Stewart, E.A., & Simons, R.L. (2010). Race, code of the street, and violent delinquency: A multilevel investigation of neighborhood street culture and individual norms of violence. *Criminology, 48*(2), 569-605.

Stewart, E.B., Stewart, E.A., & Simons, R.L. (2007). The effect of neighborhood context on the college aspirations of African American adolescents. *American Educational Research Journal, 44*(4), 896-919.

Warren, P.Y., Cochran, J., Shields, R.T., Feldmeyer, B., Bailey, C., & Stewart, E.A. (2020). Sentencing departures and female defendants: Assessing the effects of racial and ethnic threat. *Crime and Delinquency, 66*(1), 69-92.