IN THE SUPREME COURT OF THE STATE OF WASHINGTON



STATE OF WASHINGTON,

Respondent,

v.

ALLEN EUGENE GREGORY,

Appellant.

NO. 88086-7

FINDINGS AND REPORT RELATING TO PARTIES' EXPERT REPORTS

The appellant Allen Eugene Gregory proffered an updated version of a report in support of his contentions titled, *The Role of Race in Washington State Capital Sentencing, 1981 - 2014 (Updated Report)*, authored by Katherine Beckett, Professor, Law, Societies and Justice Program of the Department of Sociology at the University of Washington, and Heather Evans, M.A., Ph.D. Candidate, Department of Sociology, University of Washington. At oral argument, the State requested an opportunity to challenge the *Updated Report*. A majority of the court granted the State's request and ordered that a hearing should be held before me. The parties were directed to file memoranda addressing the conduct of the hearing, the manner of submitting testimony or other evidence, and whether the court should appoint an expert pursuant to ER 706 or alternatively, the appropriateness of appointment of a technical advisor to assist the court in understanding the evidence. Each party filed a memorandum that included suggested procedures for the State to obtain information relating to the *Updated Report's* method of analysis and conclusions, submission of additional information and evidence, and the State's presentation of the bases for its

challenge to the *Updated Report* and Mr. Gregory's response. Upon consideration of these memoranda, this court issued an order that included a provision directing the parties' attorneys to confer and determine whether agreement could be reached on the steps and timing of procedures for the following: (a) the State to obtain information relating to the *Updated Report's* method of analysis and conclusions; (b) the submission of additional information and evidence; and (c) the State's presentation of the bases for its challenge to the *Updated Report* and Mr. Gregory's responses. The parties were directed to report to me any areas where they agreed and any areas where they disagreed as to the steps and timing of such procedures.

On May 19, 2016, the parties jointly submitted an agreed proposal that consisted of the following procedures:

1. Mr. Gregory will provide the coding manual and data file for the study on the role of race in capital sentencing in Washington to the State and the Court by May 27, 2016, or within 5 days of the Commissioner's ruling detailing procedures, whichever is later.

2. By July 11, 2016, or within 45 days of receiving the data and codebook (whichever is later), the State will submit its expert report stating its conclusions about the methodology used and the reliability of the study's conclusions.

3. By August 25, 2016, or within 45 days of receiving the State's report (whichever is later), Mr. Gregory will provide the response of Professor Beckett and Ms. Evans to the State's report.

The parties did not report any areas where they disagreed as to the steps and timing of procedures. Accordingly, I accepted the parties' agreed proposal and ordered the parties to serve on the other party and file in this court the identified documents on the dates established in the agreed proposal. I also noted that I would issue a ruling detailing whether and how a technical advisor would be appointed and used.

Mr. Gregory provided the coding manual and data file to the State. The State filed the report of its expert, Nicholas Scurich, Ph.D., dated July 7, 2016, titled *Evaluation of "The Role of Race in Washington State Capital Sentencing, 1981-2014"* (*Evaluation of the Updated Report*). Mr. Gregory then submitted the response by

Professor Beckett and Ms. Evans, dated August 25, 2016, titled *Response to* "Evaluation of 'The Role of Race in Washington State Capital Sentencing, 1981-2014'" by Nicholas Scurich (Response to Evaluation).¹

In accordance with this court's order, I reviewed these documents to determine if the assistance of a neutral technical advisor with specialized skills would be beneficial. After close study of the information and explanations contained in these reports, supplemented by the Federal Judicial Center's Reference Manual on Scientific Evidence (3d ed. 2011) (Reference Manual)² and other sources, I concluded that the information available allows for the understanding of the concepts necessary to give full consideration to the experts' respective positions. However, mindful that the potential exists for a judicial officer to be confused about the technical concepts and the relationships of the technical concepts to legal principles, I indicated that it would be useful to appoint a neutral technical advisor to review my proposed report to the court for the limited purpose of identifying any areas where the proposed report reflects misunderstanding of the multiple regression methodology or interpretation of the results and to address technical explanations relating to Dr. Scurich's inability to replicate some of Professor Beckett's findings. However, on further reflection and study. I concluded these issues could be addressed by specific questions to the parties' experts. Accordingly, I issued specific questions set forth in interrogatory form and the experts responded. Professor Beckett's response to one of the interrogatories raised additional questions regarding the inclusion or exclusion of certain cases in the regression analyses reported in the Updated Report and the Response to Evaluation.

¹ In this report, I refer to the *Updated Report*, the *Evaluation of the Updated Report*, and the *Response to Evaluation*. For readability, I refer to Professor Beckett and Ms. Evans, collectively, as "Professor Beckett."

² The Federal Judicial Center provides the *Reference Manual* in a downloadable file at <u>https://www.fjc.gov/content/reference-manual-scientific-evidence-third-edition-1</u> (last visited November 20, 2017).

Accordingly, I posed additional specific questions in interrogatory form. I have now considered the answers to these interrogatories. The following findings and report are made after consideration of the Updated Report, the Evaluation of the Updated Report, the Response to Evaluation, Dr. Scurich's Answers to the Commissioner's Interrogatories, Professor Beckett's Response to Commissioner's Interrogatories to Parties' Experts, and Professor Beckett's Response to Commissioner's Supplemental Interrogatories.

The contents of this report include both factual determinations and discussion of statistical methodology. Where disagreements among the experts relate to ascertainable facts (such as claims of miscoding in the data file), I present the factual determinations that I made after considering the documentary evidence and the experts' responses to my questions. However, where the disagreements among the experts relate to statistical methodology, I set forth information that is intended to facilitate the Justices' evaluation of the different positions of the experts. This information includes the different probability values (p-values) and odds ratios that result from regression models using different assumptions about whether particular cases are properly included in the analyses.

I organize the topics in a somewhat different manner than the expert reports in order to more logically discuss the concepts from the perspective of the nonexpert. The organization follows this outline:

I. ACCURACY OF DATASET AND DATA CODING

A. Trial Reports Included in the Updated Report

B. Consistency in Data Coding

C. Data Coding Entry Errors

D. Inclusion or Exclusion of Sentencing Proceedings in Regression Analyses

1. Inclusion of Sentencing Proceedings in Which the State Stipulated That Sufficient Mitigating Circumstances Merited Leniency or That It Could Not Meet Its Burden to Prove Otherwise

- 2. Exclusion of Sentencing Proceedings on the Basis of Missing Values for One of the Variables Used in the Logistic Regression
- 3. Two Sentencing Proceedings for Same Murder Conviction: Inclusion or Exclusion of Vacated First Sentencing Proceeding

II, BASIC NUMBERS AND COMPARISONS OF PERCENTAGES

- A. Table 1: Proportion of Aggravated Murder Cases with Death-Eligible Defendants in Which Death was Sought and Imposed, by County, December 1981-May 2014.
- B. Table 2: Capital Sentence Outcomes Among Death-Eligible Washington State Aggravated Murder Defendants, December 1981-May 2014, by Race of Defendant
- C. Table 3: Capital Case Outcomes Among Death-Eligible Washington State Aggravated Murder Defendants, December 1981-May 2014, by Race of Defendant and Race of Victim

III. REGRESSION ANALYSIS

- A. Probability Values (P-Values)
 - 1. Interpretation of Statistical Significance vs. Bright-Line P-Value
 - 2. Choice of One-Tailed or Two-Tailed Test
 - 3. State Expert Challenges to and Variations of Professor Beckett's Regression Analyses and Associated P-Values
- B. Small Dataset
- C. Disclosure of Analyses Conducted
- D. Use of Parsimonious Models
- E. State Expert's Testing of the Sensitivity of the Race of Defendant Effect
 - 1. Separate Examination of White vs. Black and Other-Race vs. Black Defendants
 - 2. Race of Victim

F. Interpretation of R^2 and Pseudo R^2 Measures

IV. CONCLUSION

With this background and introduction, I now submit the following as my report to the court.

I. ACCURACY OF DATASET AND DATA CODING

A. Trial Reports Included in the Updated Report

Dr. Scurich indicates that he was asked to provide a data audit to verify the accuracy of the values reported in the *Updated Report*. As to a data audit, he states that "[i]t remains to be seen" whether any aggravated murder cases are missing from the 331 trial reports filed with this court from 1981-2014. *Evaluation of the Updated Report* at 6. He concludes, "If cases are missing, it is possible that the results would materially change." *Id.* This statement is a caveat that he has not verified the inclusiveness of the trial reports. *See* Dr. Scurich's *Answers to the Commissioner's Interrogatories* at 1.

Professor Beckett's *Response to Commissioner's Interrogatories to Parties' Experts* notes that trial reports for some of the aggravated murder cases in this timeframe are missing. *Id.* at 1. She recounts the following: in November 2013, Mr. Gregory's attorneys filed a *Motion to Complete the Process of Compiling a Full Set of Aggravated Murder Reports*; the motion was denied, but several missing trial reports were filed in subsequent months; and after these filings there were still no trial reports for several cases, including three special sentencing proceedings. She indicates that no trial report was filed for Duane Bartholomew's second special sentencing proceeding following reversal of the death sentence that was reported in Trial Report 3.³ Trial reports were filed for only two of three jury sentencing proceedings

³ See Opening Brief of Appellant at 233 n.133 (referencing Ninth Circuit Court of Appeals decision that recounts case history).

relating to Mitchell Rupe's conviction, Trial Reports 7 and 31.⁴ No trial report was filed for Charles Finch's second special sentencing proceeding following reversal of the death sentence reported in Trial Report 154. Finch committed suicide after the second jury found that life in prison without the possibility of release or parole was merited, but before judgment was entered. RCW 10.95.120 requires a report to be filed within 30 days after the entry of the judgment and sentence.⁵ Professor Beckett observes that all three of these proceedings involved white defendants who received life sentences.

The Updated Report identified 86 trial reports as cases in which a death penalty notice was filed and a special sentencing proceeding occurred. Updated Report at 28. Professor Beckett reconsidered which trial reports should be included in her Response to Evaluation, and identified 82 trial reports for inclusion in the study. Response to Evaluation at 19 n.39. These 82 trial reports included one case that was not considered in the Updated Report but was added to the calculations in the Response to Evaluation. Professor Beckett explained that it had recently come to her attention that Trial Report 34a was not an addendum to the case reported in Trial Report 34, but rather a separate case involving the same defendant but murder of a different victim committed at a different time. Response to Evaluation at 13 n.30. Paul St. Pierre, a white defendant, was convicted of aggravated murder in separate trials

⁴ This court reversed the death sentence imposed by the first jury because of the erroneous admission of Rupe's gun collection in the penalty phase of his trial. *State v. Rupe*, 101 Wn.2d 664, 683 P.2d 571 (1984). *Cf.* Trial Report 7. On remand, a second jury sentenced Rupe to death, and this court affirmed. *State v. Rupe*, 108 Wn.2d 734, 743 P.2d 210, (1987). *Cf.* Trial Report 31. Subsequently, the Ninth Circuit granted a writ of habeas corpus, concluding Rupe had a due process right to present witness polygraph results as mitigating evidence. *Rupe v. Wood*, 93 F.3d 1434, 1441 (9th Cir. 1996). In a third special sentencing proceeding the jury could not reach a unanimous verdict as required for imposition of the death penalty. *See Opening Brief of Appellant* at 66 n.32 (referencing news article).

⁵ Mr. Finch's suicide is referenced in *State v. Cross*, 156 Wn.2d 580, 649 n.11, 132 P.3d 80 (2006) (C. Johnson, J., dissenting). Review of the superior court docket indicates that a second judgment and sentence was not entered prior to the order that resolved the case due to the defendant's death.

with different juries, and was sentenced by both juries to life without the possibility of release or parole. *Id*.

Although 82 cases were included in the data file, they were not all included in the regression analysis reported in Table D, titled "Impact of Case Characteristics and Defendant Race on Capital Sentencing Outcomes in Cases with Special Sentencing Proceedings, December 1981 - May 2014." *See Response to Evaluation* at 25. A note to Table D indicates that "[i]n this model, five cases (6.1%) were missing data and were therefore dropped from the analysis." *Id.* This resulted in inclusion of 77 cases in this regression analysis. *Id.* As explained below, my questions and Professor Beckett's responses ultimately led to identification of values for the missing data in three of these five cases, and these cases were included in later analyses. Also, on her own initiative, Professor Beckett determined that the race of the defendant that Trial Report 210 (Cheyenne Brown) listed as "unknown" was white, and entered the missing value for the race of the defendant and included this case in the later analyses.⁶ Thus, an additional four cases were included in the dataset that Professor Beckett used for later analyses.

Summary: The Updated Report considered all trial reports filed and made available pursuant to RCW 10.95.120 for the identified dates of December 9, 1981 - May 31, 2014, with the exception of Trial Report 34a which was previously overlooked but considered in the Response to Evaluation. Trial reports were not filed by the trial judges for all aggravated murder cases in the December 9, 1981 - May 31, 2014, date range. As to special sentencing proceedings,

⁶ Although Trial Report 210 lists the defendant's race as unknown, Professor Beckett indicates that she subsequently matched the cause number and defendant's date of birth to entries in a database provided by the Caseload Forecast Council, and these entries list Brown's race as "white." *See Response to Commissioner's Supplemental Interrogatories* at 4. Trial Report 210 indicates the defendant entered a guilty plea and, as part of the plea bargain, the prosecutor agreed to recommend a sentence of life in prison without the possibility of release or parole in a sentencing proceeding conducted before a judge.

it is known that there are no trial reports of the second special sentencing proceeding for Duane Bartholomew or the third special sentencing proceeding for Mitchell Rupe. Although a jury entered a life sentence in Charles Finch's second trial, he committed suicide before judgment was entered. The statute does not call for the filing of a trial report in this circumstance. It should be noted that although all available trial reports were included in the data file, five trial reports were not included in earlier regression analyses because the data file had missing values for variables used in the analyses. As outlined below, available information has been inserted in the data file for four of these cases and Professor Beckett has presented analyses that include these cases in her responses to my supplemental interrogatories.

B. Consistency in Data Coding

Dr. Scurich observes that there is no description in the Updated Report that measures the degree to which coding by different raters is in agreement, particularly where "the variables require a degree of subjectivity in interpretation." Evaluation of the Updated Report at 6-7. As an example, he points out that there is no description of what constitutes "extensive publicity." Id. at 7 n.4. Professor Beckett responds that measures of inter-coder reliability are needed when coders assign numeric values to qualitative or subjective phenomena, and that here "the data entry assistants were simply entering the information provided by judges on trial reports" and not making subjective judgments. Response to Evaluation at 2.

The legislature directed the trial courts to submit reports using standard questionnaires to include specific information in all cases in which a person is convicted of aggravated first degree murder. *See* RCW 10.95.120. There is no reason to question the legislature's judgment that trial courts would be consistent in evaluating characteristics of a trial such as "[w]hether there was extensive publicity concerning the case in the community." *See* RCW 10.95.120(6)(g). For most of the

questions on the trial report form, where the coders were simply recording the trial judge's answers to questions, there is no demonstrated need for estimates of inter-rater reliability.

Professor Beckett notes one exception to her general statement that the data entry assistants were simply entering the information provided by judges on trial reports. As indicated in the *Washington State Capital Sentencing Judicial Trial Report Data, 1981-2014 Codebook (Codebook)* at 47, legal counsel coded the mitigating circumstances. In her *Response to Evaluation*, Professor Beckett explains that judges entered notations of the mitigating circumstances in a number of trial reports. She determined these notations required legal expertise to interpret, and "[f]or this reason, we relied on the legal expertise of Mr. Gregory's attorneys in coding this variable." *Response to Evaluation* at 10 n.22.

This general explanation is reasonable in light of the differences in the mitigating circumstances. of aggravating circumstances and parameters RCW 10.95.020 provides a standard and exclusive list of particular aggravating circumstances, but under RCW 10.95.070 the jury (or court if a jury is waived) may consider any relevant mitigating circumstances, including defined statutory mitigating factors and unique nonstatutory factors. However, the fact that mitigating circumstances may include such individualized factors increases the prospect that the number of mitigating circumstances may be dissimilarly counted by different reviewers of the trial reports. On a broad level, this is illustrated by noting the differences in the numbers of mitigating circumstances coded by Mr. Gregory's counsel and the numbers of mitigating circumstances listed in State v. Davis, 175 Wn.2d 287, 290 P.3d 43 (2012) for the same trial reports. In two footnotes the court indicated in parentheses the numbers of mitigating circumstances presented in trial reports as assessed by some members of the court, and I have inserted in brackets the number of mitigating circumstances coded by Mr. Gregory's counsel, as follows:

Black defendants presented the following number of mitigating circumstances: TR 29(0) [0]; TR 77(5) [11]; TR 88(2) [2]; TR 119(1) [2]; TR 135(3) [5]; TR 157(1) [2]; TR 177(0) [0]; TR 180(0) [0]; TR 185(2) [2]; TR 186(1) [5]; TR 194(1) [3]; TR 216(0) [0]; TR 281(0) [0].

White defendants presented the following number of mitigating circumstances: TR $\hat{2}(1)$ [3]; TR 3(0) [0]; TR 7(4) [6]; TR 9(0) [0]; TR 15(1) [1]; TR 16A (0) [1978 case excluded from Professor Beckett's database]; TR 20(3) [5]; TR 23(0) [0]; TR 25(2) [1]; TR 26(3) [4]; TR 31(4) [6]; TR 34(1) [2]; TR 36(0) [0]; TR 39(2) [4]; TR 42(2) [2]; TR 43(2) [2]; TR 44(2) [5]; TR 45(2) [2]; TR 47(1) [1]; TR 48(1) [1]; [juvenile defendant excluded from Professor Beckett's TR 50(4) database]; TR 51(1) [1]; TR 52(4) [5]; TR 53(1) [2]; TR 56(1) [1]; TR 58(2) [3]; TR 60(3) [2]; TR 62(0) [0]; TR 63(0) [1]; TR 64(5) [7]; TR 65(5) [6]; TR 66(1) [1]; TR 73(2) [juvenile defendant excluded from Professor Beckett's database]; TR 75(1) [2]; TR 76(0) [0]; TR 86(1) [1]; TR 92(0) [1]; TR 93(1) [3]; TR 95(1) [2]; TR 125(1) [3]; TR 132(1) [2]; TR 140(2) [5]; TR 144(2) [2]; TR 154(1) TR 165(1) [2]; TR 167(1) [2]; TR 174(1) [3]; [1]; [2]: TR 164(2) 175(1) Ī1Ī: TR [1]; TR 176(3) [3]; TR 183(0) Ī0]; TR 181(0) TR = 182(1)[2];TR 184(2) [2]; [4]; TR 190(4) TR 220(5) [6]; TR 227(2) [7]; TR 251(2) [3]; TR 258(3) [4]; TR 303(3) [5].

Id. at 369-70 n.69 & n.70 (bracketed material added). Neither the *Codebook* nor the *Updated Report* articulates the basis for numbers coded by Mr. Gregory's counsel. The *Codebook* description of the coding protocol is cryptic.⁷ For example, the coders are instructed to enter "1" if the court found "one kind" of mitigating evidence, without explaining what "one kind" means. Additionally, review of the coding results suggests that the *Codebook* instructions were not always followed. Notwithstanding

⁷ The *Codebook* at 63 contains the following instruction:

- 55. Court's determination of credible evidence of mitigating circumstances: Did the court find any credible evidence of mitigating circumstances?
 - a. Enter 0 if the court found no credible evidence of mitigating circumstances.
 - b. Enter 1 if the judge lists one kind of mitigating evidence that s/he found credible.
 - c. Enter 2 if the judge lists two kinds of mitigating evidence that s/he found credible.
 - d. Enter 3 if the judge lists three or more kinds of mitigating evidence that s/he found credible.

(Emphasis in original.)

the *Codebook* instruction to "[e]nter 3 if the judge lists three or more kinds of mitigating evidence that s/he found credible," cases were coded with higher numbers where the judges listed more than three kinds of mitigating evidence. Accordingly, there is some merit to Dr. Scurich's challenge that the *Updated Report* does not provide an intelligible coding manual or information on the efficacy of coding as to the number of mitigating circumstances.

In providing background information for Interrogatory No. 2, I indicated that my review of the trial reports and the coding entries largely revealed the coding practices for mitigating circumstances were as follows: examine trial report responses to question 3(c) ("Was there, in the court's opinion, credible evidence of any mitigating circumstances as provided in Laws of 1981, ch. 138, § 7? If yes, please describe."), and question 3(d) ("Was there evidence of mitigating circumstances, whether or not of a type listed in Laws of 1981, ch. 138, § 7, not described in answer to (3)(c) above? If yes, please describe."); evaluate the number of individual concepts as separated by the trial judge in the trial report; enter the number of concepts in answer to question 3(c) on the coding sheet under "MitCircum_Statutory: Statutory mitigating circumstances" and in answer to question 3(d) on the coding sheet under "MitCircum_NonStat: Non-statutory mitigating circumstances"; and add the two numbers and enter the total in the coding sheet under "MitCircum_Total: Total mitigating circumstances." Further, in the background to Interrogatory No. 2, I stated that my review of the trial reports and coding entries indicated that Mr. Gregory's counsel, acting as coders, endeavored in most instances to simply quantify the number of individual concepts as separated by the trial judges in listing evidence of mitigating circumstances, without further evaluation. A few examples illustrate this observation. The 11 total mitigating circumstances coded for Trial Report 77 seems comparatively high. But in answer to question 3(c) the trial judge indicated there was credible evidence of four of the eight factors listed in RCW 10.95.070 that the jury may

consider in deciding if leniency is merited. In answer to question 3(d) the trial judge numbered and listed an additional seven types of mitigating circumstances with some specificity, including the following that the coder counted as two of the seven mitigating circumstances: "(2) The defendant's prior family history that would reasonably be expected to contribute to his criminal conduct. (3) The abuse and neglect of the defendant when he was a child." In Trial Report 119 the trial judge described the evidence of mitigating circumstances as "[t]estimony of friends and family members regarding the killing of his father by his mother and other aspects of his life." The coder marked this as "2" separate mitigating circumstances. In Trial Report 39 the trial judge described the evidence of mitigating circumstances as follows: "Youth had long history of treatment as an abused family. Father had sexually abused defendant's sister. Had mental problems." The coder entered "4" for the total number of mitigating circumstances.⁸

In her answer to Interrogatory No. 2, Professor Beckett confirmed that I correctly identified the general coding protocol for the mitigating circumstances variable and provided more detail, stating:

Specifically, the protocol was as follows:

1. Enter the number of discrete statutory mitigating circumstances from question 3(c) into the Mit_Circum_Statutory field.

Professor Beckett's Response to Commissioner's Interrogatories to Parties' Experts at 2.

⁸ The background to Interrogatory No. 2 questioned whether there were instances where coders exercised judgment rather than recording the number of concepts reported by the trial judge, but in answering this interrogatory Professor Beckett provided the following information:

The Commissioner correctly identified one data entry error: The number of mitigating circumstances for TR 25 should have been 2, not 1. In light of this discovery, we re-checked *all* trial reports for cases that had a special sentencing proceeding, and confirmed that TR 25 was the only one with such an error. We did, however, discover that a 0 rather than a 1 had been entered for the number of aggravating circumstances found by the judge/jury for the proceeding associated with TR 34A, so also corrected this.

- a. If the "No" box is checked, enter zero.
- b. If the "Yes" box is checked with no description, enter one.
- c. If nothing is checked and the description field is empty, enter zero.
- d. If nothing is checked but the judge has described one or more mitigating circumstances, count them and enter the appropriate number.
- e. If one or more mitigating circumstances in this field is actually non-statutory, count it in the Mit Circum Non-Stat field instead.
- 2. Under question 3(d), evaluate the number of individual concepts described by the trial judge and enter that number into the Mit_Circum_Non-Stat field.
 - a. If the "No" box is checked, enter zero.
 - b. If the "Yes" box is checked with no description, enter one.
 - c. If nothing is checked and the description field is empty, enter zero.
 - d. If nothing is checked but the judge has described one or more mitigating circumstances, count them and enter the appropriate number.
 - e. If one or more descriptions in this field is actually statutory, count it in the Mit Circum Statutory field instead.
- 3. Add the numbers from 3(c) and 3(d) and enter the total into the TotMitCircum field.

As these protocols illustrate, a value of "0" is coded for mitigating circumstances in a variety of situations. At one end of the range is Trial Report 177, where the trial judge marked "No" in response to both questions as to whether there was credible evidence of any mitigating circumstances and further indicated that "[t]he defendant requested the jury to impose the death penalty and would not assist defense counsel in presenting mitigation evidence contrary to his desires." At the other end of the range is Trial Report 23 where the trial judge did not mark either "yes" or "no" under the questions as to whether there was evidence of mitigating circumstances and left blank the lines provided for descriptions of the evidence, and the coder entered "0" mitigating circumstances.

The protocol descriptions do not indicate how to code the number of mitigating circumstances if the trial report indicates that the defendant and the prosecutor stipulated in a special sentencing proceeding before a judge that there were sufficient mitigating circumstances to merit leniency, as in Trial Reports 92 and 182,

or stipulated that the State could not prove beyond a reasonable doubt that there were not sufficient mitigating circumstances to merit leniency, as in Trial Reports 167 and 224.9 Comparison of the trial reports and the data file entries shows three of these cases were coded as "1" mitigating circumstance: Trial Report 92, where the trial judge checked "ves" and indicated that "the defendant & prosecutor stipulate that there was mitigating circumstances;" Trial Report 167, where the trial judge did not check either "yes" or "no" but wrote "the defendant and the State stipulated that the State could not prove beyond a reasonable doubt that there were not sufficient mitigating circumstances to merit leniency;" and Trial Report 224, where the trial judge wrote "[n]ot applicable" and then explained that "[i]n lieu of evidence at the special sentencing proceeding, the state and the defendant submitted a stipulation, accepted by the court, 'that the State of Washington is not able to prove beyond a reasonable doubt that there are not sufficient mitigating circumstances to merit leniency." Trial Report 182 was coded as "2" mitigating circumstances; the trial judge checked "yes" under question 3(c) and stated that "[t]he court accepted the parties' stipulation that mitigation was sufficient for leniency" and checked "yes" under question 3(d) and described the mitigating circumstances as follows: "Childhood abuse resulting in psychological disorder. Pled guilty as charged without knowing whether the Prosecutor would continue to pursue the death penalty at the special sentencing proceeding." Though not originally explained, coding of mitigating circumstances has now been described and, as a general matter, is internally consistent.

While Dr. Scurich raised broad concerns about inter-coder reliability, neither he nor Professor Beckett specifically addressed the coding of aggravating

⁹ While background to Interrogatory No. 8 stated that it appeared these cases were not included in the regression analyses, Professor Beckett indicates that "TRs 92, 167, 182 and 224 were included in all of the regression analyses." Professor Beckett's *Response to Commissioner's Interrogatories to Parties' Experts* at 5 (emphasis in original). The inclusion of these cases in the regression analyses is discussed in Section I.D.1.

circumstances that were found to be applicable. Inconsistencies in coding for this independent variable first came to my attention as I reviewed Professor Beckett's answers to my supplemental questions relating to multiple sentencing proceedings. See, infra, notes 40 and 41 (discussing apparent inconsistencies in the coding of the number of aggravating circumstances for the first and second sentencing proceedings for two defendants). A broad look at the data file suggests coders without legal expertise made entries that varied with the format of the trial judge's entry on the trial report, even where the substance of the reported information was the same. Coders were instructed to "enter the number of aggravating circumstances found by the judge to have been applicable in this case, e.g., 1." Codebook at 62. This number was entered in the data file column EZ, "AppliedAggCir Num." See id. at 45. The trial judge reported information on applicable aggravating circumstances in response to trial report form question 2(e), "What aggravating circumstances, as set forth in Laws of 1981, ch. 138 § 2, were alleged against the defendant and which of these circumstances were found to have been applicable?" Two columns follow, the first with the subheading "Aggravating Circumstances Alleged" and blank lines below, and the second with the subheading "Found Applicable" with boxes for checking "Yes" or "No" as to the alleged aggravating circumstance. It came to my attention that different numbers of aggravating circumstance were entered for trial reports in which the same statutory aggravating circumstances set forth in RCW 10.95.020(9)-(11) were found applicable. These aggravating circumstances include the following:

(9) The person committed the murder to conceal the commission of a crime or to protect or conceal the identity of any person committing a crime, including, but specifically not limited to, any attempt to avoid prosecution as a persistent offender as defined in RCW 9.94A.030; (10) There was more than one victim and the murders were part of a common scheme or plan or the result of a single act of the person; (11) The murder was committed in the course of, in furtherance of, or in immediate flight from one of the following crimes:

- (a) Robbery in the first or second degree;
- (b) Rape in the first or second degree;
- (c) Burglary in the first or second degree or residential burglary;
- (d) Kidnapping in the first degree; or
- (e) Arson in the first degree;

RCW 10.95.020. By way of example, Trial Report 180 (Cecil Davis) lists (on one line) the crimes in RCW 10.95.020(11)(a), (b), and (c) without citation to the statute, followed by a checked "yes" box. Trial Report 281 (Cecil Davis) cites "RCW 10.95.020(11)(a), (b), (c)" and lists on the next line the three particular crimes in those subsections, with a single checked "yes" box opposite the first line. For Trial Report 180 the coder entered "1" aggravating circumstance and for Trial Report 281 the coder entered "2" aggravating circumstances. I perused other trial reports to see how reports of applicable subsections of RCW 10.95.020(11) were counted, and noted that in Trial Report 140 (Cal Brown) each subsection was counted as an aggravating circumstance. Trial Report 140 lists each of the following on separate lines, with each followed by a checked "yes" box: "Rape 1° & 2°," "Robbery 1° & 2°," "Kidnapping 1°," and "Conceal Identity." The coder entered "4" aggravating circumstances in the data file, obviously counting each subsection of RCW 10.95.020(11) as one aggravating factor. Additionally, I noted inconsistencies in coding aggravating circumstances in cases involving multiple victims. Some trial reports indicate the aggravating circumstances found applicable without setting forth the number of counts, and other trial reports list the aggravating circumstances found applicable as to each count. The coding of the number of aggravating circumstances depended on this format difference. The following examples illustrate this type of coding inconsistency. Trial Report 13 (Kwan Mak) indicates that the death sentence was imposed as to 13 counts of first degree murder, but does not refer to the multiple counts when it lists the aggravating circumstances found applicable. The aggravating circumstances found applicable are listed as "murder committed to conceal crime or identity" and "murder in course of robbery." The coder entered "2" aggravating circumstances in the data file. In Trial Report 43 (David Rice) the trial judge wrote "AS TO EACH COUNT" before listing on three lines the aggravating circumstances found applicable, "10.95.020(7)," "10.95.020(8)," and "10.95.020(9)."¹⁰ The coder entered "12" aggravating circumstances, apparently multiplying the number of aggravating circumstances by the four counts of first degree murder. In Trial Report 31 (Mitchell Rupe), the same three aggravating circumstances were listed by the name of the crimes in response to question 2(e), but without a reference on that part of the form to the two counts of first degree murder. The coder entered "3" aggravating circumstances in the data file.¹¹

Summary: In light of the standard trial report form that is completed by judges, Dr. Scurich's concerns relating to inter-coder reliability are inapplicable to most of the coding entries since the coders recorded trial judge report answers without application of evaluative criteria. However, these concerns were potentially applicable to the coding of mitigating circumstances by counsel, since the criteria to be used by the coders was not initially described. Professor Beckett has now described the coding protocols. Even before this description was provided, a comparison of the trial reports and the coding sheet showed that the general approach of the coders was to record the number of individual concepts as separated by the trial judges in listing evidence of mitigating circumstances. Review shows that this approach was generally followed in an even-handed manner. However, Professor Beckett's description of the protocols does not describe the coding procedure to be used where the trial report indicates that the parties stipulated that mitigating circumstances merited leniency or that the State could not meet its burden to prove

¹⁰ These subsections are now codified as RCW 10.95.020(9), (10), and (11).

¹¹ For additional examples, *compare* Trial Reports 9, 39, 53, 76, 132, 144 (coding the number of aggravating circumstances by including each aggravating circumstance found to be applicable as to each count), *with* Trial Reports 14, 15, 20, 86, 174 (coding the number of aggravating circumstances found to have been applicable without regard to whether the aggravating circumstance applied to more than one count).

there were not sufficient mitigating circumstances to merit leniency. Review of trial reports and coding shows that where such a trial report contained no further description of mitigating circumstances, the trial report was coded as "1" mitigating circumstance. In one such instance, the trial report described two concepts of mitigating circumstances, and the trial report was coded as "2" mitigating circumstances. Coding of mitigating circumstances has now been explained or can be discerned and is internally consistent. However, it appears there are inconsistencies in the coding of aggravating circumstances found applicable. These inconsistencies likely stem from a combination of various formats used by trial judges to report aggravating circumstances and coding by persons without the expertise necessary to understand the substance of the information reported.

C. Data Coding Entry Errors

Dr. Scurich identified three coding errors that Professor Beckett agreed should be corrected. The three errors and the corrections that were made are summarized as follows:

- (1) Jack Owen Spillman (Trial Report 167) was incorrectly coded as having "received the death penalty" where the trial report indicates that he was sentenced to life in prison without the possibility of release or parole. He was recoded as receiving a life sentence.
- (2) Gary Michael Benn (Trial Report 75) was incorrectly coded as an "other race" defendant. The trial report lists his race or ethnic origin as "Caucasian." He was recoded as "white."
- (3) Richard Blake Pirtle (Trial Report 132) was incorrectly coded as "did not receive a death sentence" where the trial report states that a sentence of death was imposed. He was recoded as having received a death sentence.

See Evaluation of Updated Report at 26 and Response to Evaluation at 14. Additionally, in his initial evaluation, Dr. Scurich appeared to suggest that a coding error treated a black defendant as a white defendant in the analysis of whether black defendants are more likely than nonblack defendants to receive the death penalty.¹² As to the three coding errors identified in Dr. Scurich's report, none treated a black defendant as a white defendant. The only coding error related to race was the miscoding of Benn as an "other race" defendant when he should have been coded as "white." The original codings of Spillman as "white" and Pirtle as "white" were correct. In an answer to Interrogatory No. 3, Dr. Scurich clarified that he is not aware of evidence that a coding error treated a black defendant as a white defendant in the analysis of whether black defendants are more likely than nonblack defendants to receive the death penalty, explaining that the statement was intended simply to illustrate how easily and unwittingly a coding error could occur. Dr. Scurich indicates, "When I stated '[t]here is evidence that this actually occurred . . .' it would have been more appropriate to state '[t]here is evidence that this *type of coding error* actually occurred."" Dr. Scurich's *Answers to the Commissioner's Interrogatories* at 1 (emphasis in original).

In light of the identified coding errors, and Dr. Scurich's statement that he had "no idea how often such errors occurred in the current data file," *Evaluation of the Updated Report* at 6 n.3, I conducted my own review by looking at the trial report descriptions of the defendants' race on trial reports coded as cases in which a death penalty notice was filed and a special sentencing proceeding was conducted. My purpose was to determine if there were any coding errors as to the defendants' race in

¹² Dr. Scurich made the following suggestion in the course of discussing how coding errors can alter results: "For instance, if black defendants are to be coded as '2', it is possible that an occasional error could cause a black defendant to be coded as '1', the code for a white defendant. Such a mistake would go undetected unless every single variable code for every single case were independently verified (and even then such errors can get overlooked). However, such an error could completely alter the results, in that it treats a black defendant as a white defendant in the data analysis. *There is evidence that this actually occurred in the analysis predicting whether blacks are more likely than non-blacks to receive the death penalty (see section 2.4)*. Since there was no attempt to estimate inter-rater reliability (consistency), we simply have no idea how often such errors occurred in the current data file." *Evaluation of the Updated Report* at 6 n.3 (emphasis added).

these cases. Additionally, Dr. Scurich noted that defendant race was classified as "white, black or other race" and observed that there was no explanation of how potentially ambiguous cases were handled. *Id.* at 85. To consider this challenge, I reviewed the information in the trial reports as to the race of defendants classified as "other race."

In the following trial reports where death notices were filed and the case proceeded to a special sentencing proceeding, the defendant is listed as "black" or "African American" (except where variations of these terms are otherwise noted): 29, 77,¹³ 88, 119, 135, 157, 177, 180, 185, 186, 194, 216, 281, 312. These 14 trial reports match the denominator of 14 black defendants in Table A of Professor Beckett's *Response to Evaluation* at 16. The *Codebook*, at 58, instructs the use of the code "2" for defendant's race if recorded in the trial report as "black or African American." I checked each of the 14 trial reports and each was coded "2" under "D_Race."

In the following trial reports where death notices were filed and the case proceeded to a special sentencing proceeding, the defendant is listed as "Caucasian" or "white" (except where variations of these terms is otherwise noted): 2, 3, 7, 9,¹⁴ 15, 20, 23, 25, 26,¹⁵ 31, 34, 34a, 36, 39, 42,¹⁶ 43, 44, 45, 47, 48, 51, 52, 53, 56, 58, 62, 63, 64, 65, 66, 75, 76, 86, 92, 93, 95, 125, 132, 140, 144, 154, 164, 165, 167, 174, 175, 176, 182, 183, 184, 190, 220, 227, 251, 258, 303, 313. These 57 trial reports match the denominator of 57 white defendants in Table A, *Response to Evaluation* at 16. The *Codebook*, at 58, instructs the use of the code "1" for defendant's race if recorded in the trial report as "white or Caucasian." I checked each of these 57 trial reports and

¹³ Listed in trial report a "Black (father black, mother caucasian)."

¹⁴ Listed in trial report as "white (some Hawaiian ancestry)." In answering a question about the percentage of the county population that is the same race as the defendant, the trial judge indicated as follows: "Notwithstanding the trace of Hawaiian descent, I would regard Mr. Campbell as being of the white race."

¹⁵ Listed in trial report as "appears to be Caucasian."

¹⁶ Listed in trial report as "European/N. American."

each was coded "1" under "D_Race," with two exceptions. As noted above, Dr. Scurich identified the original miscoding of Trial Report 75 as "other race." Professor Beckett states this miscoding has been corrected in the analyses presented in the *Response to Evaluation* and later analyses. Also, Trial Report 34a was not included in the original data file and I therefore could not check the coding.

In the following trial reports where death notices were filed and the case proceeded to a special sentencing proceeding, the trial reports list the defendants as "other race," with the trial judge statement as to race appearing as noted: 8,¹⁷ 13,¹⁸ 14,¹⁹ 60,²⁰ 158,²¹ 160,²² 181,²³ 197,²⁴ 224,²⁵ 256.²⁶ These ten trial reports match the denominator of 10 "other race" defendants in Table A, *Response to Evaluation* at 16. The *Codebook*, at 58-59, instructs the use of the code "3" for Hispanic or Latino/a; "4" for Native American or Alaskan Native; "5" for Asian or Pacific Islander; and "6" for "other race." I checked each of these 10 trial reports and each was coded 3, 4, 5 or 6 under D Race.

My review shows that in the group of 57 included in the "white" category the trial judge described the defendant as "Caucasian" or "white" in 54 cases and in the other three cases describe the defendant variously as "appears to be white," "white

¹⁷ Listed in trial report as "¼ to ½ Ponca Indian & caucasi[a]n."

¹⁸ Listed in trial report as "Asian."

¹⁹ Listed in trial report as "Asian."

²⁰ Listed in trial report as "½ Caucasi[a]n & ½ American Indian."

²¹ Listed in trial report as "Native."

²² The section of Trial Report 160 (Jeremy Sagastegui) asking the race or ethnic origin of the defendant is blank, but a different section of the trial report states that "[t]he three victims were Caucasian" and further indicates the victims were not the same race or ethnic origin as the defendant. The State's brief in another case, *State v. Thomas*, 150 Wn.2d 821, 83 P.3d 970 (2004), in listing the race of persons executed since the death penalty was reinstituted, included "Jeremy Sagastegui (Hispanic)." *See* bound volumes of briefs, Wash. State Law Library, 150 Wn.2nd 821 Briefs, Vol. 11; *Brief of Respondent* at 164.

²³ Listed in trial report as "Caucasian/Native American."

²⁴ Listed in trial report as "Native American."

²⁵ Listed in trial report as "Latino (Mexican)."

²⁶ Listed in trial report as "Caucasian/Asian."

with some Hawaiian ancestry"²⁷ and "European/N. American." In the 14 cases included in the "black" category, the trial judge described the defendant as "black" or "African American" in 13 of the cases; in the remaining case the trial judge described the defendant as "Black (father black, mother caucasian)." Of the ten cases Professor Beckett placed in the "other race" category, the trial judge described the defendant as "Asian" in two cases, "Caucasian/Asian" in one case, Caucasian/Native American in three cases, Native American in two cases, and "Latino (Mexican)" in one case. In the tenth case, the place on the form to indicate race or ethnic origin was left blank but the trial judge indicated that the defendant was of a different race from the white victims and the State's counsel has previously represented that this defendant was Hispanic. *See* note 22, *supra*.

In sum, I reviewed all trial reports where the coding indicated that death notices were filed and the case proceeded to a special sentencing proceeding and found only Trial Report 75 was miscoded as to the defendant's race. In answer to interrogatories, Professor Beckett confirmed that the trial reports I have listed correspond to the denominators of 14 black defendants, 57 white defendants, and 10 "other race" defendants in Table A of the *Response to Evaluation* at 16. *See* Professor Beckett's *Response to Commissioner's Interrogatories to Parties' Experts* at 4.

Summary: Professor Beckett acknowledges and corrects coding errors identified by Dr. Scurich, including one coding error as to the race of the defendant and two coding errors as to the sentence that was imposed. Additionally, Professor Beckett has identified and corrected two coding or data entry errors related to the number of mitigating circumstances and number of aggravating circumstances found by the judge or the jury. Review of the cases included in Professor Beckett's revised group of "cases in which death notices were filed and not withdrawn" as set forth in

²⁷ As noted, the trial judge stated that notwithstanding a "trace" of Hawaiian ancestry, he considered the defendant's race as white.

her Response to Evaluation reveals no basis to question the placement of these cases in the assigned racial categories.

D. Inclusion or Exclusion of Sentencing Proceedings in Regression Analyses

1. Inclusion of Sentencing Proceedings in Which the State Stipulated That Sufficient Mitigating Circumstances Merited Leniency or That It Could Not Meet Its Burden to Prove Otherwise

In responding to Dr. Scurich's *Evaluation of the Updated Report*, Professor Beckett adjusted the cases that she included in the calculations and analyses related to sentencing proceedings. Professor Beckett excluded Trial Reports 152 and 153, explaining "defendants subsequently entered a stipulated guilty plea and a special sentencing hearing therefore did not occur." *Response to Evaluation* at 16 n.37. This reasoning is similar to the explanation in the *Codebook* that Trial Report 81 "was not included in the analysis of jury decision-making because a plea deal was reached before the special sentencing proceeding." *Codebook* at 4. Trial Report 81 indicates that the parties reached a plea agreement in which the prosecutor agreed to recommend life without the possibility of parole, and notes the date on which a special sentencing proceeding was held before the judge. Although Trial Reports 152 and 153 provide specific dates on which a special sentencing proceeding commenced, Professor Beckett maintains that review of the trial reports as a whole suggests that the death notice had been withdrawn. *See* Professor Beckett's *Response to Commissioner's Interrogatories to Parties' Experts* at 5.

As I observed in background accompanying Interrogatories No. 7 and No. 8, Trial Reports 92, 167, 182, and 224, are also cases in which the defendant and the prosecutor reached agreements and the special sentencing proceeding was conducted before a judge. Trial Report 92 indicates "the proceeding was conducted before a judge" and that "the defendant & prosecutor stipulate that there was mitigating circumstances." Trial Report 167 states that the defendant pleaded guilty and "[w]ith the consent of the State, he then waived jury for the special sentencing proceeding," and further indicates that "the defendant and the State stipulated that the State could not prove beyond a reasonable doubt that there were not sufficient mitigating circumstances to merit leniency." Trial Report 182 states that "[t]he jury for special sentencing was waived by the parties and the court accepted the waiver" and further indicates, "[t]he court accepted the parties' stipulation that mitigation was sufficient for leniency." Trial Report 224 states that "[t]he defendant pled guilty and waived his right to a jury trial for the special sentencing proceeding," and that "[i]n lieu of evidence at the special sentencing proceeding, the state and the defendant submitted a stipulation, accepted by the court, 'that the State of Washington is not able to prove beyond a reasonable doubt that there are not sufficient mitigating circumstances to merit leniency." I asked Professor Beckett whether Trial Reports 92, 167, 182, and 224 were included in the calculation of percentages of cases in which death notices were filed and juries imposed death sentences across racial groups. Additionally, I indicated that my review suggested that Trial Reports 92, 167, 182 and 224 may have been the four cases excluded from the dataset of special proceedings in the regression analysis reported in Table D in the Response to Evaluation at 25, and asked whether they were in fact included in this analysis.²⁸ As discussed, Professor Beckett indicated that these four cases were included in both the percentage calculations and the analysis. See Professor Beckett's Response to Commissioner's regression Interrogatories to Parties' Experts at 5.

Professor Beckett states that she included Trial Reports 92, 167, 182, and 224 in the regression analyses for the following reasons:

²⁸ Table D shows that the model included 77 cases. *Response to Evaluation* at 25. This is 4 fewer cases than Professor Beckett's amended dataset of 81 cases with special sentencing proceedings. *Cf. id.* at 16, note to Table A.

We included all special sentencing proceedings in our analyses of sentencing decisions because if a special sentencing proceeding occurred, a sentence of death could have been imposed by the judge or jury. The difference between the included and excluded cases is that in the former, a special sentencing proceeding actually occurred. Although it is true that in three of these four cases the prosecution stipulated during this proceeding that it could not prove beyond a reasonable doubt that there were not sufficient mitigating circumstances to warrant leniency, the death notices do not appear to have been withdrawn, special sentencing proceedings actually took place, and the judge, exercising independent judgment based upon the facts of the crime, could still have imposed a death sentence.

By contrast, in the excluded cases (TRs 81, 152 and 153), there was no evidence that a special sentencing proceeding actually occurred. Although the judge did enter a date of a special sentencing proceeding, it is clear that this was simply the date on which the sentence was imposed; no information about a special sentencing proceeding is provided. Instead, these trial reports indicate that there was a guilty plea with an agreement of a life without parole sentence, suggesting that the death notice had been withdrawn. It thus appears that in these cases, the defendant was sentenced for non-capital aggravated murder after offering a guilty plea based on an agreement of a life without parole sentence. In these cases, the judge was precluded from sentencing the defendant to death as a result of the structure of the plea.deal. The judge would not have had the legal ability to weigh the facts of the crime and determine whether death was appropriate. For this reason, these three cases were not included in our analyses of sentencing outcomes.

Response to Commissioner's Interrogatories to Parties' Experts at 5-6 (footnote observing that the three excluded proceedings involved white defendants who killed multiple victims omitted). Professor Beckett also indicates that Table A in the *Response to Evaluation* at 16 should be retitled *Percent of Special Sentencing Proceedings in which a Death Sentence was Imposed, by Race of Defendant* (rather than *Percent of Aggravated Murder Cases with Special Sentencing Proceedings in which Juries Imposed a Death Sentence by Race of Defendant*).²⁹ Id. at 8.

Professor Beckett's conclusion that "the judge, exercising independent judgment based upon the facts of the crime, could still have imposed a death

²⁹ Professor Beckett points out that the *Updated Report* at 5 n.19, indicated as follows: "If a defendant waives his or her right to a jury trial, a judge may impose a death sentence in cases in which a death notice has been filed. As a practical matter, however, juries almost always decide whether to impose a sentence of death. We therefore link sentencing decisions to jury decision-making throughout this article."

sentence" is inconsistent with this court's description of these cases in *Davis*, 175 Wn.2d 287. In *Davis* the court observed that in the cases reported in Trial Reports 92, 167 and 182 three white defendants received a life sentence after the prosecution stipulated that mitigating circumstances merited leniency, *id.* at 366, and noted that Trial Report 224 reported similar circumstances in a case involving a Mexican defendant. *Id.* at 366 n.49. The court concluded as to these cases, "The life sentences imposed in these cases obviously had nothing to do with the judge's or jurors' attitudes about race. Indeed, by stipulating that mitigating evidence merited leniency, *id.* at 366.

Summary: The stated purpose of the regression analyses of sentencing decisions is to determine if the race of the defendant impacted the decision maker, whether a judge or a jury. In Trial Reports 92, 167, 182, and 224 the defendant and the prosecutor stipulated that there were sufficient mitigating circumstances to merit leniency or that the prosecutor could not meet the State's burden to prove there were not sufficient mitigating circumstances to merit leniency. If the prosecutors' stipulations in Trial Reports 92, 167, 182, and 224 gave the decision makers no alternative as to the sentence imposed, as indicated by this court's decision in Davis, these cases were improperly included in the regression analyses.³⁰

2. Exclusion of Sentencing Proceedings on the Basis of Missing Values for One of the Variables Used in the Logistic Regression

Professor Beckett's response that these four cases were included in the regression analyses led to the question of which four cases were excluded. While Professor Beckett did not identify cases that were excluded from various regression

³⁰ As noted above, Professor Beckett on her own initiative entered a missing value for the race of the defendant for Trial Report 210 and included this case in her analyses reported in her *Response to Commissioner's Supplemental Interrogatories*. *Id.* at 4. Trial Report 210 indicates the defendant entered a guilty plea in exchange for the prosecutor's agreement to recommend a sentence of life in prison without the possibility of release or parole, with the special sentencing proceeding conducted before a judge. There is no mention in this trial report of an evidentiary stipulation.

analyses of sentencing decisions by trial report or name, she did explain that the number of cases used in an analysis may be smaller than the total number of cases if there are missing values for any of the variables used in the logistic regression. *Response to Evaluation* at 55. For the model reported in Table D, "the dataset analyzed includes 77 special sentencing proceedings with no missing values." Professor Beckett's *Response to Commissioner's Interrogatories to Parties' Experts* at 44. A note to Table D indicated that "[i]n this model, five cases (6.1%) were missing data and were therefore dropped from the analysis." *Response to Evaluation* at 25.

After receiving Professor Beckett's responses to my first set of questions, I reviewed the data file to ascertain which cases were missing values for a variable used in the analyses reported in Table 7 of the Updated Report and Table D of the Response to Evaluation. The variables used were the number of prior convictions, one victim, number of applied aggravators, number of mitigating circumstances, number of defenses, whether a victim was held hostage, and black defendant. In answers to various interrogatories, Professor Beckett indicated that two cases were missing information on the number of prior convictions, one case was missing information on the number of defenses, and one case was missing information on whether the victim See Professor Beckett's Response to Commissioner's held hostage. was Interrogatories to Parties' Experts at 7, 32, 28 n.15. One case in which the race of the defendant was listed as "unknown" (Trial Report 210) had been previously excluded from the analyses reported in the Updated Report and Response to Evaluation.³¹ See Updated Report at 21 and Response to Evaluation at 61. Professor Beckett indicated that no cases were missing information on the number of victims or the number of aggravating circumstances found. Id. at 28 n.15. If information is missing for a

³¹ As discussed above, Professor Beckett later ascertained the race of the defendant through another judicial agency source and added this value to the data file, such that later analyses refer to 82 cases.

variable that is included in an analysis, the software program automatically excludes the case from the analysis. Cf. Response to Evaluation at 4 n.9 (explaining this feature of the software in a different context). In my review I found that the data file column "AR" with the heading "D Priors," where information on the number of prior convictions is entered, was blank only for Trial Report 8 (Charles Bingham) and Trial Report 15 (Patrick Jeffries). The data file column for the number of defenses, "EP" under the heading "Defenses Num," was blank only for Trial Report 313 (Byron Scherf). Finally, I looked to see which trial report had missing data on whether a victim was held hostage. The data file column "DT" under the heading "Vics AnyHostage" was blank only for Trial Report 197 (Joseph Revay). In a supplemental interrogatory, I set forth this review and asked Professor Beckett whether the five cases that were dropped from the analyses reported in the Response to Evaluation were Trial Reports 210, 8, 15, 313 and 197. In response, Professor Beckett confirmed that these cases were dropped from previous analyses due to missing values in the data file. Response to Commissioner's Supplemental Interrogatories at 1.

However, as I indicated in background to the supplemental interrogatories, my review of Trial Reports 8, 15, and 313 showed these reports actually contained the information that the data file indicated was missing.³² Both trial reports indicate on the trial report form that prior convictions are listed on an attachment. Mr. Gregory's counsel responded to my supplemental questions on this point. They stated that although Trial Reports 8 and 15 each indicated that prior convictions were listed on an

³² The trial report form instructs at question 1(i), "If the defendant has a record of prior convictions, please list." Immediately under the form instruction, Trial Report 8 (Charles Bingham) states, "See excerpt from Bill of Particulars filed in this cause by the Prosecuting Attorney, which lists the entire record of the Defendant's prior convictions, attached hereto." The attached bill of particulars details 8 prior convictions. Similarly, in Trial Report 15 (Patrick Jeffries) immediately under the form instruction the trial judge wrote, "See copy of Canadian record of convictions attached hereto as Exhibit A." The attached exhibit lists 15 prior convictions.

attachment, the copies of the trial reports provided to counsel did not include the attachments and informal requests to the clerk's office for copies of the attachments were not fruitful. As described below, Professor Beckett has now inserted the number of prior convictions³³ in the data file and has provided the results of analyses that include these cases. As to Trial Report 313, where the value for the number of defenses was left blank in the data file, Professor Beckett indicates "this was an error" and a value of "0" has been inserted in the data file such that the case has been included in subsequent analyses.³⁴

Summary: Trial Report 8 (Charles Bingham), Trial Report 15 (Patrick Jeffries), Trial Report 313 (Byron Scherf), and Trial Report 197 (Joseph Revay) were not included in the regression analyses related to the "Impact of Case Characteristics and Defendant Race on Capital Sentencing Outcomes in Cases with Special Sentencing Proceedings, December 1981 - May 2014," the results of which are reported in Table 7 of the Updated Report and Table D of the Response to Evaluation.

³³ Mr. Gregory's counsel indicated they still had not seen the attachments listing the prior convictions, and raised the possibility that the number of prior convictions indicated in the background to the supplemental interrogatories were erroneous, noting that I indicated that Trial Report 8 listed 8 convictions and Trial Report 15 listed 15 convictions. *Response to Commissioner's Supplemental Interrogatories* at 2. I have double checked, and do not see any error; the matching numbers are a coincidence. At my request, on October 11, 2017, the clerk of the court forwarded copies of the attachments to the parties.

³⁴ In background to the supplemental interrogatories, I noted that the trial report form question 2(c) instructs, "Please indicate if there was evidence introduced or instructions given as to any defense(s) to the crime of aggravated first degree murder" followed by a list of specific defenses and aligned boxes that can be checked to indicate "Evidence" and "Instruction(s)." The trial report form also provides blank lines for "[o]ther specific defenses" to be written in by the judge, with accompanying boxes for checking. The Codebook instructs coders to "[e]nter 0 for no" and "[e]nter 1 for yes" as to whether evidence was given regarding a specific defense or "other specific defense" and to enter the total number of defenses for which evidence was entered. Codebook at 61. The values for total number of defenses range from 0 to 4. See id. at 43. All trial reports for cases with special sentencing proceedings had a value of "0" to "4" entered in the data file column "EP" under the heading "Defenses_Num" except Trial Report 313. In Trial Report 313 the trial judge wrote "general denial" on one of the blank lines under the form subheading "Other specific defenses," but did not check any box. Professor Beckett states that "[b]ecause neither the evidence nor the instructions box was checked for the sole defense identified in TR 313, the correct number should be zero." Response to Commissioner's Supplemental Interrogatories at 3.

Each of these cases was excluded on the basis that there were missing values for one of the variables used in the logistic regression. The exclusion of Trial Reports 8 and 15 is attributable to blanks in the data file because attachments to the trial reports were not available to coders. These blanks have been filled by entry of the number of prior convictions that were listed on trial report attachments referenced in the trial reports. As to Trial Report 313, the trial judge made a handwritten entry "general denial" in response to question 2(c) ("Please indicate if there was evidence introduced or instructions given as to any defense(s) to the crime of aggravated first degree murder"). Professor Beckett has stated that it was error to leave the data file blank and that a value of "0" has now been entered. The analyses reported in Professor Beckett's Response to Commissioner's Supplemental Interrogatories include Trial Report 8 (Charles Bingham), Trial Report 15 (Patrick Jeffries), and Trial Report 313 (Byron Scherf). I did not pose any supplemental questions regarding Trial Report 210 (Cheyenne Brown) that listed the defendant's race as "unknown" or ask that it be included in any analyses. On her own initiative Professor Beckett noted other judicial agency records listed Brown's race as white, added this value to the data file, and included this case in all of the later analyses. No information is available relating to the missing value in Trial Report 197, and it was not included in any analyses.

3. Two Sentencing Proceedings for Same Murder Conviction: Inclusion or Exclusion of Vacated First Sentencing Proceeding

Dr. Scurich observes that the regression analyses in the Updated Report included two cases for Allen Gregory (Trial Reports 216 and 312), Mitchell Rupe (Trial Reports 7 and 31), and Cecil Davis (Trial Reports 180 and 281). As to each of these defendants, the first trial report relates to a special sentencing proceeding in which a death sentence was imposed that was later reversed and the second trial report relates to a subsequent special sentencing proceeding. Dr. Scurich opines that inclusion of such "redundant" observations violates the assumption of independence in logistic regression, which he describes as follows:

A central assumption of logistic regression is that each observation is independent. As a leading text on multivariate statistics put it:

'Logistic regression assumes that responses of different cases are independent of each other. That is, it is assumed that each response comes from a different, unrelated case.'

This assumption is plainly violated when one individual contributes multiple 'cases' to the datafile.

Evaluation of the Updated Report at 25 (quoting Barbara G. Tabachnick & Linda S. Fidell, *USING MULTIVARIATE STATISTICS* at 445 (6th ed. 2013) (footnotes and page reference omitted). In a footnote, Dr. Scurich further quotes Tabachnick & Fidell as stating that "[t]he effect of non-independence in logistic regression is to produce overdispersion," and that "this results in an inflated Type I error rate^[35] for tests of predictors." *Evaluation of the Updated Report* at 25 n.21 (quoting *id.* at 445). Dr. Scurich states, "In short, the reported p-value will be an underestimate of the 'true' p-value." *Id.* On this basis, Dr. Scurich asserts that it is inappropriate for Professor Beckett to include in the model both the original jury proceeding in which a death sentence was imposed and, following reversal, the second sentencing proceeding for the same defendant. He contends that only the second sentencing proceeding should be included. *Evaluation of the Updated Report* at 25.

Professor Beckett generally agrees with the principle that cases examined in logistic regression models should be independent one from the other, but takes the position that the assumption of independence is not violated where the two sentencing proceedings were decided by different juries and involved some different case

³⁵ The "Reference Guide on Statistics" in the *Reference Manual* defines a Type I error as follows: "A statistical test makes a Type I error when (1) the null hypothesis is true and (2) the test rejects the null hypothesis, i.e., there is a false positive. . . . When a statistical test deems the difference to be significant in this situation, it makes a Type I error." *Id.* at 300-01. Type I (false positive) and Type II (false negative) errors are discussed at some length in the expert reports.

characteristics. She argues that in these circumstances the assumption of

independence should not be broadly applied, stating as follows:

It is true that three defendants in the dataset had second trials and that these second trials are included in our analyses. This is appropriate because in our study, the unit of analysis is the outcome (specifically, the decision to file a death notice or impose a death sentence), not the defendant. The three cases that Dr. Scurich removed from one of his model tests because he believed them to be "redundant" involved defendants (including Mr. Gregory) who had second trials that involved newly constituted juries and different case characteristics, and therefore could very well have resulted in a different outcome. In Mr. Gregory's case, for example, many such differences between the two trials exist, including the number of mitigating circumstances and the number of prior convictions. In addition, the two trials were separated by eleven years and involved different juries and defense attorneys. Dr. Scurich suggests that including both trials violates the assumption that the cases included in the regression model are independent, an assumption upon which regression analyses theoretically depend (see pp. 25-27 of his critique). While one can argue that a defendant's second trial is not entirely independent of his or her first trial, the argument can also be made that any trials involving the same judges, prosecutors, or defense attorneys are also not entirely independent of each other. In fact, if one interprets the assumption of independence broadly, cases adjudicated by the same judge, or in the same county, could be said to violate the assumption of independence. Given the very significant differences that can characterize the three second trials from the first trials, and the fact that the second juries plainly could have made a different sentencing decision, we believe it is most appropriate to include both trials in the dataset - while also remembering that regression results are always and inevitably mathematical estimates of real-world processes.

Response to Evaluation at 10-11 (emphasis in original) (footnotes omitted). Professor

Beckett reiterates this point in her Response to Commissioner's Supplemental

Interrogatories stating as follows:

[A]lthough it is true that a defendant's second special sentencing proceeding is not entirely independent of his or her first proceeding, it is also true that any proceedings adjudicated by the same judge, involving the same attorneys, or adjudicated in the same county violate the assumption of independence. Nonetheless, as a practical matter, researchers using regression methods to analyze sentencing outcomes routinely include cases that involve the same judges, attorneys, counties, and defendants in their analyses, and the results are often published in well-regarded, peer-reviewed journals. *Id.* at 7 (footnote with journal citations omitted.) Professor Beckett also notes differences in the number of prior convictions as to two of the defendants, differences in the number of mitigating circumstances as to Mr. Gregory's first and second trials, and differences related to the passage of time and the involvement of different attorneys. *See id.* at 10.

There are sources of support for each expert's position. Dr. Scurich's position—that inclusion of both sentencing proceedings when the death sentence is reversed and the defendant is sentenced a second time violates the assumption of independence—is consistent with the conclusions of Judge Baime, a special master who reported to the New Jersey Supreme Court on proportionality review of death penalty sentencing. Judge Baime retained statisticians Dr. David Weisburd, director of the Institute of Criminology at The Hebrew University, and Dr. Joseph Naus, a professor of statistics at Rutgers University, to advise him as he considered the cases that should be included in multiple regression analyses of death penalty sentencing outcomes. These experts recommended including only one disposition in cases where the defendant has been tried multiple times for the same murder, explaining their concern as follows:

It does not make statistical or substantive sense to count the exact same case multiple times. This problem is exacerbated by the fact that in this situation one of the cases will by definition always be a death outcome (i.e. the first case). Many of the statistical tests employed in analyses we propose, as was the case with earlier statistical analyses conducted by the AOC [Administrative Office of the Courts], assume that the cases examined are independent one from another. This assumption is seriously violated when we include the same murder case tried multiple times in a single analysis.

Professor David Weisburd & Professor Joseph Naus, Report to Special Master David Baime (Technical Appendix) 16, attached to David S. Baime, Report to the New Jersey Supreme Court: Systemic Proportionality Review Project 2000-2001 Term (June 1, 2001) (Baime Report III). Judge Baime described this aspect of independence as follows: "Multiple regression rests on the assumption that there is no systematic relationship between measured characteristics of the cases and unmeasured characteristics that influence death penalty sentencing. This assumption may or may not be reasonable for cases involving different individuals, but it is certainly suspect when there are multiple cases involving the same individual in a single analysis." *Baime Report III* at 27. The "Reference Guide on Statistics" in the *Reference Manual* provides some insight into the basis for this position. Certain assumptions are made about the "error term," defined as "[t]he part of a statistical model that describes random error, i.e., the impact of chance factors unrelated to variables in the model." *Id.* at 287. In such models, "[t]he errors are assumed to be independent and identically distributed from person to person in the dataset. Such assumptions are critical when computing *p*-values and demonstrating statistical significance." *Id.* at 281. Accordingly, if there is less confidence that each of the two observations contributes independent information to the regression analysis, there is less confidence in the calculation of the magnitude of the random error.³⁶

In support of her position, Professor Beckett cites articles in peer reviewed publications where Washington State sentencing data were analyzed from datasets that included cases that involved the same judges, attorneys, and counties. None of these articles discusses use of multiple sentencing proceedings related to the same

³⁶ Take, by way of example, a regression model that includes independent variables for diet and exercise habits and results in an odds ratio that individuals who eat high fat diets are three times as likely as individuals who eat low fat diets to contract a certain disease. A reported low p-value of .03 (a 3 percent probability of getting, just by chance, a test statistic as large as or larger than the observed value) will underestimate the probability of random error if the study includes multiple siblings such that there is a systematic relationship between shared but unmeasured genetic factors.

defendant and crime in regression analyses.³⁷ However, my own research into this issue found support for Professor Beckett's focus on the independence of the decision makers in the two proceedings. *See* David Baldus, George Woodworth, David Zuckerman, Neil Alan Weiner & Catherine M. Grosso, *Empirical Studies of Race and Geographic Discrimination in the Administration of the Death Penalty: A Primer on the Key Methodological Issues*, in THE FUTURE OF AMERICA'S DEATH PENALTY: AN AGENDA FOR THE NEXT GENERATION OF CAPITAL PUNISHMENT RESEARCH 153 (Charles S. Lanier, William J. Bowers & James R. Acker eds., 2009). These authors note that an independence issue exists when there are subsequent penalty proceedings, but they focus on the independence of the decision makers, as follows:

[I]n studies with sufficient data to model penalty trial outcomes, an independence issue arises with respect to the inclusion of subsequent penalty trial sentencing decisions in the analysis. In contrast to prosecutorial decision making, subsequent penalty trial decisions are reasonably treated as independent of previous decisions unless there is evidence that a subsequent jury was told that the defendant had earlier been sentenced to death. However, if a subsequent prosecution advances to a bench trial, the judge would be aware of the earlier decision, suggesting that such a case should be treated like a subsequent prosecutorial charging decision and coded accordingly.

Id. at 159. This approach appears to apply a concept of independence similar to the definition supplied by the "Reference Guide on Statistics" in the *Reference Manual* at 288, which defines "independence" as follows: "Events are independent when the

³⁷ Articles that Professor Beckett cites as examples where Washington State sentencing data were analyzed, *Response to Commissioner's Supplemental Interrogatories* at 7 n.5, include the following: Randy R. Gainey, Sara Steen and Rodney L. Engen, "Exercising Options: An Assessment of the Use of Alternative Sanctions for Drug Offenders", *Justice Quarterly* 22:4, 488-520 (2005); Rodney L. Engen, Randy R. Gainey, Robert D. Crutchfield, and Joseph G. Weis, "Discretion and Disparity Under Sentencing Guidelines," *Criminology* 41, 1: 99-130 (2003); Rodney L. Engen, "The Power to Punish: Discretion and Sentencing Reform in the War on Drugs," *American Journal of Sociology* 105, 5: 1357-1395 (2000); and Alexes Harris, Heather Evans and Katherine Beckett, "Courtesy Stigma and Monetary Sanctions: Toward a Socio-Cultural Theory of Punishment," *American Sociological Review* 76, 2: 234-64 (2011). These articles do not discuss the use of multiple sentences related to the same defendant and crime in regression analyses. Only the first article mentions the assumption of independent observations, explaining steps taken to ensure certain estimates were independent of county characteristics. *See Justice Quarterly* 22:4 at 503.
probability of one is unaffected by the occurrence or non-occurrence of the other."³⁸ Baldus, Woodworth, Zuckerman, Weiner & Grosso do not address the concept of "error term" or the extent to which aspects of the first and second trials were different or similar.

There is some judgment involved in assessing the degree to which similarities or differences in cases involving the same defendant are substantial enough to cause a concern about the independence of the observations. This point is illustrated by the reasons the technical experts gave Judge Baime for concluding that it is reasonable to include all cases of a defendant who committed multiple murders on different occasions. They explained as follows: "While such cases involve the same defendant, the victims and circumstances of the cases are different in each case. We are still concerned with the lack of independence of the cases due to the involvement of the same offender, but the number of cases here is not very large and the substantive differences in other characteristics of the cases are substantial enough to convince us that the statistical analyses we conduct will not be strongly affected." Professor David Weisburd & Professor Joseph Naus, *Report to Special Master David Baime*, (Technical Appendix) 16, attached to David S. Baime, Report to the New Jersey Supreme Court: Systemic Proportionality Review Project 2000-2001 Term (June 1, 2001) (*Baime Report III*).³⁹

³⁸ Baldus, Woodworth, Zuckerman, Weiner, & Grosso contrast the independence of subsequent penalty trial decisions with decisions to bring a second prosecution. Where a death sentence is vacated and subsequent decisions are made to again prosecute, the authors observe that the first prosecution may influence decisions in subsequent prosecutions, and "discretionary judgments made in the earlier case may simply be repeated in the later cases(s)" and thus reflect prosecutorial decisions that are not independent. Baldus, *supra.*, at 158.

³⁹ I note that the New Jersey Supreme Court's proportionality review project initially included a single case for each defendant for statistical purposes, even where the same defendant committed murders on separate occasions, based on the experts' earlier concern about the assumption of independence. *See In re Proportionality Review Project (II)*, 165 N.J. 206, 219-20, 757 A.2d 168 (2000). The special master and experts subsequently changed the recommended approach in these circumstances, as reflected in this later report.

With this background, the question here is whether differences in the characteristics of the first and second sentencing proceedings convince one that there were no factors unaccounted for in the regression model that substantially affected the death penalty sentencing decisions of both the first and second juries. The answer to this question will identify whether there is concern about the independence of the observations if both the first and second sentencing proceedings are included in a regression analysis. As noted, this is a question of judgment and one appropriately left to the Justices.

Professor Beckett also questions the basis for choosing which of the two sentencing proceedings to include in the regression model. She adds, "Moreover, there is no rational basis for deciding which of a defendant's two special sentencing proceedings to include. Deciding to keep the first but exclude the second, or vice versa, would be arbitrary, but because the proceedings involve different characteristics, either decision could have an impact on the results." Professor Beckett's Responses to Commissioner's Interrogatories to Parties' Experts at 10. But it appears to be generally accepted by statisticians that if inclusion of two cases would violate the assumption of independence, the answer is to run alternative models that show the results if one selects one set of cases or another. As noted above, when the special master for the New Jersey Supreme Court and his advising experts found no convincing basis to choose one proceeding over another, they ran a "first case sample" which included only the first case for each defendant who was tried multiple times for the same murder, and a "last case sample" that included the second disposition for each of these defendants. Cf. Baldus, supra, at 158 n.10 (recognizing selection among a "first" prosecution decision and a "last" prosecution decision for regression analyses as one of several alternatives that could address the lack of independence in multiple prosecution decisions).

Dr. Scurich included only the second sentencing proceeding for each of these defendants in his models. Although he does not state his reasons for including the second sentencing proceedings rather than the first, use of the second sentencing proceedings appears appropriate. I note that the decisions reversing the first sentencing proceeding for each of these three defendants find that the jury may have been influenced by improper factors. *See State v. Rupe*, 101 Wn.2d 664, 704-07, 683 P.2d 571 (1984) (reversed death sentence because the trial court erroneously admitted evidence of Rupe's gun collection in the penalty phase of his trial); *In re Pers. Restraint of Davis*, 152 Wn.2d 647, 704-05, 101 P.3d 1 (2004) (granting personal restraint petition and reversing sentence because jurors had seen Davis in shackles); *State v. Gregory*, 158 Wn.2d 759, 777-78, 147 P.3d 1201 (2006) (reversing death sentence because the prosecutor improperly commented that there were positive aspects of prison living conditions).

While another approach would have been to run alternative "first case sample" and "second case sample" analyses, it does not appear there would be a significant difference in the results of the samples. The values shown in the data file for the variables used in these regression analyses were the same for Rupe's first and second sentencing proceedings. There are some differences in the values for Mr. Gregory's first and second sentencing proceedings, but the differences appear minimal for purposes of the results of the regression model. The number of prior convictions for Mr. Gregory's first proceeding was 12 and for the second proceeding 9. The number of prior convictions is not shown as predictive of sentencing outcomes in any of the regression analyses results reported in the tables in Professor Beckett's *Response to Commissioner's Supplemental Interrogatories*. The number of mitigating circumstances entered for Mr. Gregory's first proceeding was 0 and for the second proceeding 1. The regression analyses results reported in the *Response to*

Commissioner's Supplemental Interrogatories indicate that if adding a mitigating circumstance affects a sentencing outcome, it decreases the likelihood of a death sentence, such that use of the second trial report would attribute more of the likelihood estimate to the race of the defendant. (Although Mr. Gregory's first proceeding was coded as 2 aggravating circumstances and the second proceeding was coded as 1 aggravating circumstance, the values should have been the same, but the coders evidently arrived at different numbers regarding the number of aggravating circumstances in the two proceedings because the same information was presented in two ways.⁴⁰) As to Davis, 9 was entered as the number of prior convictions in the first proceeding and 10 was entered as the number of prior convictions in the second proceeding. (Davis's first proceeding was coded as 1 aggravating circumstance and the second proceeding was coded as 2 aggravating circumstances, the trial reports clearly reported the same information and the values should have been the same.⁴¹)

⁴⁰ Trial Report 216, in response to question 2(e) regarding what aggravating factors set forth in the statute were alleged and which were found to have been applicable, lists "[RCW] 10.95.020(11)(b)-Murder Committed in course of Rape 1 or Rape 2" followed by a checked box that the circumstance was found applicable, and then lists "[RCW] 10.95.020(1[1])(a)-Murder Committed in course of Robbery 1" followed by a checked box that the circumstance was found applicable. In the same section of the trial report form, Trial Report 312 lists "RAPE 1, 2 (RCW 10.95.020(11)(b)); ROBBERY 1 (RCW 10.95.020(11)(a)" with a single box checked. The first proceeding was coded as "2" aggravating circumstances and the second proceeding was coded as "1" aggravating circumstance in column EZ, for the AppliedAggCir__Num (Number of aggravating circumstances found by the judge to have been applicable in this case.) *See Codebook* at 45. Professor Beckett evidently recognized that aggravating factors were in fact the same, as she did not identify any differences between the first and second proceedings as to this value even though the coding differed.

⁴¹ Trial Report 180, in response to question 2(e) regarding which aggravating factors set forth in the statute were alleged and which were found to have been applicable, lists "Murder committed during: Robbery 1/2, Rape 1/2, Burglary 1/2" followed by a single checked "yes" box under the column "Found Applicable." Trial Report 281 lists "RCW 10.95.020(11)(a), (b), (c) Robbery 1, 2; Rape 1, 2; Burglary 1, 2" followed by a single checked "yes" box under the column "Found Applicable." The first proceeding was coded as "1" aggravating circumstance and the second proceeding was coded as "2" aggravating circumstances. Again, Professor Beckett evidently recognized that aggravating factors were in fact the same; she does not identify any differences between the first and second proceedings as to this value even though the coding differed.

Summary: Inclusion of two sentencing proceedings in the dataset of a logistic regression model, where the first death sentence was reversed and the defendant was sentenced a second time for the same murder, raises questions about whether regression analyses that include two sentencing proceedings for the same defendant meet the central assumption of logistic regression that each observation is independent. Under this independence assumption, only one sentencing proceeding is appropriately included in a dataset for regression analysis if factors not measured in the regression model substantially influenced the decisions of both the first and second juries. If such circumstances exist here, it is appropriate to include only one of the sentencing proceedings in a regression model. If only one proceeding is included, Dr. Scurich's inclusion of the second sentencing proceedings appears to be appropriate since errors in the first sentencing proceeding identified improper evidence or observations that could have made a difference in the sentencing decisions. Another approach would be to run alternative "first case sample" and "last case sample" analyses. But there is little indication the regression results would differ significantly if data file values for the first proceedings were used instead of the values for the second proceedings.⁴²

II. BASIC NUMBERS AND COMPARISONS OF PERCENTAGES

When conducting the data audit, Dr. Scurich copied seven tables from the Updated Report and indicated whether he was able or unable to verify the

⁴² Because Trial Report 34a was added to the dataset after Dr. Scurich's evaluation of the *Updated Report*, he did not have an opportunity to express an opinion on whether inclusion of both Trial Report 34 and Trial Report 34a comports with the assumption of logistic regression that each observation is independent. I asked Professor Beckett to provide certain calculations with and without the addition of Trial Report 34a so that both calculations would be available if an issue arose. (I did not ask Dr. Scurich for a calculation with the inclusion of Trial Report 34a because the data file he was provided does not include the coding for this case.) If the views of the New Jersey proportionality project experts on this point are accepted as persuasive, inclusion of Trial Report 34a is appropriate.

information. Since Tables 4-7 reflect the results of regression models, I consider comments on these tables under the heading for regression analysis below. In this section, I consider Dr. Scurich's *Evaluation of the Updated Report* as to Tables 1-3, which involve numbers and comparison of percentages without the application of regression models.

A. <u>Table 1: Proportion of Aggravated Murder Cases with Death-Eligible Defendants</u> in Which Death was Sought and Imposed, by County, December 1981-May 2014.

Dr. Scurich questions the logic of the column "Proportion of Aggravated Murder Cases in which Death Penalty was Imposed" in Table 1, where the denominator 297 includes cases in which no death notice was filed. *Id* at 8. *Cf. Updated Report* at 20. He indicates that the denominator should be 86, the number of death notices that were filed. Professor Beckett responds that the use of each of these denominators generates a different measure. She states that use of 297 aggravated murder cases as the denominator is "intended to provide readers with a broad sense of county-level variation in the share of aggravated murder convictions that resulted in a death sentence in Washington State – regardless of the precise mechanism that explained this variation." *Response to Evaluation* at 14

The purpose of Professor Beckett's use of 297 as the denominator can be inferred from its placement in Table 1 of the *Updated Report* since it appears alongside the number of cases in which death notices were filed (i.e., 86/297). This purpose is made explicit in the *Response to Evaluation* at 14. However, the description of Table 1 as showing the "share of aggravated murder *convictions* that resulted in a death sentence" in each county, *id*. (emphasis added), is not wholly accurate. The numbers include multiple proceedings for the same defendant whether or not the conviction for the crime was vacated. For example, in Thurston County there was one conviction and two sentencing proceedings for the same crime, with the

death penalty imposed in both sentencing proceedings. Table 1 indicates that in Thurston County the death penalty was imposed in "33% (2/6)," Updated Report at 20, whereas the numerator relates to only one conviction and the denominator relates to 5 convictions. In response to Interrogatory No. 10 Professor Beckett notes that the title and discussion indicate the unit of analysis is each "case" or each special sentencing proceeding, not conviction, and that she inadvertently used the word "conviction" in the penultimate sentence when discussing these findings, *Response to Evaluation* at 14, when she should have used the word "case." With this clarification, she indicates the numbers are correct.

Dr. Scurich also states he was not able to verify the numbers in the last two columns in Table 1. First, he states "I was not able to verify the numbers in the 'average number of victims' column. This variable does not appear in the datafile or the codebook. It is also not explicitly defined in the Report, leaving it unclear as to what the average refers to exactly (e.g., average number of victims per defendant, per case, etc.)." Evaluation of the Updated Report at 9 (emphasis in original). Professor Beckett explains in the Response to Evaluation at 15 that this average number was not the value of a variable; rather, an average was calculated by summing the number of victims in each case and dividing that sum by the number of cases in a particular county then rounding to the nearest whole number. Dr. Scurich next indicates, "I was not able to verify the numbers in the 'average number of affirmed aggravators' column" because a variable with this description does not appear in the data file or the codebook. Evaluation of the Updated Report at 9 (emphasis in original). Professor Beckett clarifies that "we consider the number of aggravated circumstances found by the jury to be applicable to be affirmed aggravators." Response to Evaluation at 15 n.34.

Dr. Scurich notes that not all values reported in Table 1 are correct, and shows that the data file includes 53 trial reports for Pierce County with the total number of victims 82, and that 82 divided by 53 is 1.547, which rounds to 2, not 3. *Answers to the Commissioner's Interrogatories* at 2-3. My review indicates Dr. Scurich is correct on this point.

Summary: To the extent Dr. Scurich was not able to verify the numbers in the last two columns in Table 1 because of uncertainty about the descriptions, Professor Beckett has provided clarification of the descriptions. Dr. Scurich is correct when he points out that the value reported in Table I for the average number of victims for Pierce County is incorrect, and should be 2 rather than 3.

B. <u>Table 2: Capital Sentence Outcomes Among Death-Eligible Washington State</u> <u>Aggravated Murder Defendants, December 1981 - May 2014, by Race of</u> <u>Defendant</u>

As to Table 2, Dr. Scurich first notes that the total numerator "86" in the "All" row exceeds by one the sum of the numerators for the number of death notices filed as to defendants in three race categories: white, black, and other race. Table 2 is accompanied by the following: "Note: Defendant race is unknown in one case." *Updated Report* at 21. Professor Beckett mistakenly added the "unknown" case to the numerator but not to the denominator in the "death notice filed" column of Table 2, but this resulted in no difference to the percentage calculation.

Next, Dr. Scurich states the denominators in the "death penalty imposed" column are incorrect because they represent the total number of aggravated murder cases and he believes "[t]he appropriate denominator is 86 (the number of cases in which a death notice was filed), not the total number of cases (296), since the death penalty cannot be imposed if a death notice is not filed." *Evaluation of the Updated Report* at 10. Professor Beckett responds that the intent of Table 2 is to provide a view

of the percentages that reach the designated category from the broad group of "death-eligible" cases and therefore use of a denominator associated with all cases is appropriate. The careful reader would understand, and Professor Beckett has now clarified, that the denominators represent numbers and percentages related to all aggravated murder cases.

In narrative following Table 2 of the Updated Report, Professor Beckett calculated and compared the percentages of cases in which death notices were filed and juries imposed death sentences across racial groups. The Updated Report coded 86 cases as having a death notice filed. One of these cases was removed from the calculations on the basis the trial report lists the race of the defendant as "unknown." Professor Beckett then calculated percentages using the remaining 85 cases. She concluded that "we can calculate that juries imposed death in 37% of the cases involving white defendants, but 64% of the cases involving black defendants, in which prosecutors filed a death notice." Updated Report at 21.

Professor Beckett reconsiders which cases should be included for purposes of this calculation in the *Response to Evaluation*. She observes that in five of the 85 cases included in the *Updated Report*, intervening events prevented the jury from considering death as a sentence. *See Response to Evaluation* at 16 n.37. In three of the five cases, Trial Reports 68, 217, and 308, prosecutors filed death sentences but the defendants were later ruled ineligible for special sentencing proceedings.⁴³ *Id*. Professor Beckett indicates that two additional cases were removed because plea agreements "took death sentences off the table." *Id*. at 16, note to Table A. She indicates that in Trial Reports 152 and 153 "defendants subsequently entered a

⁴³ These cases (unlike other excluded cases listed in the *Codebook*) were among the cases that were coded. These trial reports were coded "0" or "No" for "Death penalty sought" and evidently were included in the original percentage calculations. The *Codebook* at 3, indicates these cases "were not included in the analysis because although death notices were filed by prosecutors, legal rulings prevented a special sentencing proceeding."

stipulated guilty plea and a special sentencing hearing therefore did not occur."⁴⁴ *Id.* at 16 n.37. As noted above, Professor Beckett also added one case, noting that "[i]t recently came to our attention that Trial Report 34A was not simply an addendum, but rather contained information about a separate case involving the same defendant described in Trial Report 34. This case was not included in the analyses presented in our report. The defendant in question is Paul St. Pierre, a white man who was convicted of two separate aggravated murders and was sentenced by both of his juries to life without the possibility of parole." *Id.* at 13 n.30. These 81 cases (85-5+1 = 81) are the cases that comprise the denominators (14 for black defendants and 67 for nonblack defendants) used in Table A found in the *Response to Evaluation* at 16.

Additionally, Professor Beckett's recalculation of the percentages adjusts the numerators and denominators of the "white defendants" and "other race defendants" to account for the coding error as to Trial Report 75 discussed previously, in which the defendant was coded as "other race" when the trial report describes the defendant's race as "Caucasian." Correction of this error subtracts one from the numerator and denominator of the "other race" category and adds one to the numerator and denominator of the "white" category. The *Updated Report* indicated that a jury imposed a death sentence in 22 of 60 cases with white defendants, but Professor Beckett's *Response to Evaluation* indicates that a jury imposed the death sentence in 23 of 57 cases with white defendants. And while the *Updated Report*

⁴⁴ Professor Beckett indicates the table in Appendix C of the *Updated Report* shows "cases in which death notices were filed and special sentencing hearings occurred." *See Response to Evaluation* at 16. Trial Report 152 indicates there was a "guilty plea per stipulation" and Trial Report 153 indicates "guilty plea made by stipulation." Dates were given for the special sentencing proceedings, and in the blank following the question as to the jury's findings, the trial judge cross-referenced the guilty pleas by stipulation. (The *Codebook* at 4 indicates that TR 81 had previously been excluded from the analysis of jury decision-making because a plea deal was reached before the special sentencing proceeding.)

defendants, Professor Beckett's *Response to Evaluation* at 16 indicates that a jury imposed the death sentence in 3 of 10 cases with "other race" defendants.

The narrative that follows Table A states that this table shows that "in cases in which death notices were filed and not withdrawn, *juries imposed death in 38.8 percent of the cases involving death eligible non-Black defendants, but 64.3 percent of otherwise similar cases involving Black defendants*. The racial gap between White and Black defendants in [sic] nearly as large." *Id.* (emphasis in original). (The percentage for "white defendants" was 23 out of 57 cases, or 40.4% and for "other race" the percentage was 3 out of 10 cases or 30%.)

While the *Response to Evaluation* at 16 states that the percentages reflect cases in which *juries* imposed a death sentence, this characterization is inaccurate as to the denominators. It is evident that the denominators include four cases that did not involve juries at the special sentencing proceedings. The "other race" denominator includes Trial Report 224, which indicates the defendant pleaded guilty and waived his right to a jury trial for the special sentencing proceeding and that the State and the defendant submitted a stipulation, accepted by the court, that the State was not able to prove beyond a reasonable doubt that there were not sufficient mitigating circumstances to merit leniency. The "white" denominator includes three cases in which the trial reports indicate that sentencing proceedings were not held before a jury: Trial Reports 92, 167,⁴⁵ and 182.⁴⁶ As noted above, in responses to

⁴⁵ A coding error related to Trial Report 167 did not affect these percentages. As discussed, the defendant was miscoded as having received the death penalty when he did not, but another white defendant was miscoded as having received a life sentence without the possibility of release when in fact he received the death penalty. Accordingly, these coding errors "wash out" for purposes of the calculation of basic percentages.

interrogatories, Professor Beckett clarifies that the denominator cases include four or five cases with special sentencing proceedings before a judge rather than a jury, and that the correct title of the table is *Percent of Special Sentencing Proceedings in which a Death Sentence was Imposed by Race of Defendant* (rather than *Percent of Aggravated Murder Cases with Special Sentencing Proceedings in which Juries Imposed a Death Sentence by Race of Defendant*).⁴⁷ *Response to Commissioner's Interrogatories to Parties' Experts* at 8.

As discussed above, if in stipulating that mitigating evidence merited leniency, the prosecution gave the judges in Trial Reports 92, 167, 182 and 224 no alternative but to impose a sentence of life without the possibility of parole, it is arguable whether these cases should be included in the percentages. Calculations show slightly different percentages if these cases are omitted. If Trial Report 224 is removed from the denominator for "other race" defendants, juries or judges imposed death in 33.3% of cases involving death eligible "other race" defendants. If three cases are removed from the denominator for white defendants (Trial Reports 92, 167 and 182), juries or judges imposed death in 42.6% of cases involving death eligible

⁴⁶ Additionally, although Trial Report 63 was filled out as if sentencing proceedings were held before a jury, the defendant Leslie McVay was tried to the court sitting without a jury and it was the court that imposed a sentence of life without parole. In completing the trial report, the trial judge evidently confused the circumstances of the trial with that of the defendant in Trial Report 62, Thomas Kron. The different proceedings to which each were subject are described in *State v. Kron*, 63 Wn. App. 688, 691, 821 P.2d 1248 (1992). Kron solicited McVay to kill his former girlfriend. Kron was convicted by a jury of aggravated first degree murder and the State sought the death penalty. He was sentenced by the jury to a life sentence without parole. McVay was charged jointly with Kron, but his case was severed from Kron's and was tried first to the court sitting without a jury. A superior court judge found McVay guilty on all counts and imposed a sentence of life without parole, but sealed the penalty until Kron's trial was finished. The same judge also presided over Kron's jury trial. There is some indication the trial judge had in mind the different defendants in completing the trial reports since the answers as to whether there was evidence of mitigating circumstances was different in the two cases.

⁴⁷ Professor Beckett points out that the *Updated Report* at 5 n.19, indicated as follows: "If a defendant waives his or her right to a jury trial, a judge may impose a death sentence in cases in which a death notice has been filed. As a practical matter, however, juries almost always decide whether to impose a sentence of death. We therefore link sentencing decisions to jury decision-making throughout this article."

white defendants. If these groups are combined into one group of "nonblack" defendants, removal of these four cases from the combined denominator results in a calculation that juries or judges imposed death in 41.3% of cases involving death eligible "nonblack" defendants. There are no changes to the calculation that juries or judges imposed death in 64.3% of cases involving death eligible black defendants.

Table 2 also contains a column with the heading "Death Penalty Retained." Updated Report at 21. Dr. Scurich reports: "I was not able to verify the numbers in the 'death penalty retained' column, since this variable does not appear in the data file or the codebook." Evaluation of the Updated Report at 10 (emphasis in original). The Updated Report at 21 explains the term "retained" in a note relating to Table 2, as follows: "Retained' in this context means that the death sentence was not reversed by a higher court or was re-imposed after reversal of the original death sentence." Standing by itself, and in light of the ordinary meaning of "retained" that suggests action to hold or to keep, I would have interpreted this use of "retained" as including only cases in which an appeal from the sentencing decision was decided. But Professor Beckett's response to Interrogatory No. 15 indicates that the "death penalty retained" column of Table 2 includes direct appeals of the death sentences that are pending, including Trial Report 303 (Schierman) and Trial Report 216 (Gregory).48 She indicates that Trial Report 313 (Scherf) should have been included, but "Scherf (TR 313) was incorrectly coded as not having the death penalty retained so was excluded from this table." Professor Beckett's Response to Commissioner's

⁴⁸ Presumably, Professor Beckett intended to refer to Trial Report 312 since this court reversed the death sentence that is reported in Trial Report 216.

Interrogatories to Parties' Experts at 9.⁴⁹ Accordingly, while Table 2 lists eight white defendants as to whom the death penalty was "retained," Professor Beckett would now include nine white defendants in the retained category, which she has identified by trial report number and name. These nine cases include the following, with information that I have added in brackets:

TR 9 (Campbell) [executed]

TR 76 (Dodd) [executed]

TR 140 (Brown) [executed]

TR 165 (Elmore) [death sentence affirmed on appeal; personal restraint petition dismissed and federal habeas relief denied]

TR 183 (Elledge) [executed]

TR 220 (Cross) [death sentence affirmed on appeal; personal restraint petition dismissed; federal habeas petition pending]

TR 251 (Yates) [death sentence affirmed on appeal; personal restraint petition dismissed; federal habeas petition pending]

TR 303 (Schierman) [direct appeal pending]

TR 313 (Scherf) [direct appeal pending]

Table 2 lists one "other race" defendant as to whom the death penalty was "retained,"

which refers to the following defendant:

TR 160 (Sagastegui) [executed].

Table 2 lists four black defendants as to whom the death penalty was "retained." In answer to Interrogatory No. 15, Professor Beckett indicates that she included the cases in the following trial reports, with information that I have added in brackets:

⁴⁹ Professor Beckett indicates that "'retained status' is not included in any of the regression analyses, so the regression results are also unaffected by the omission of Scherf (TR 313) from Table 3 of the *Updated Report*." *Response to Commissioner's Interrogatories to Parties' Experts* at 9. However, as I outlined above, the software program omitted Trial Report 313 from the regression analyses of sentencing decisions reported in the *Updated Report* and *Response to Evaluation* because of a coding error. Professor Beckett corrected this coding error and indicates this trial report was included in the analyses reported in her *Response to Commissioner's Supplemental Interrogatories*.

TR 119 (Gentry) [death sentence affirmed on appeal, two personal restraint petitions dismissed or denied; federal habeas relief denied; third personal restraint petition pending]

TR 177 (Woods) [death sentence affirmed on appeal, personal restraint petition denied; application for federal habeas relief dismissed in January 2017 due to Woods's death]

TR 180 [likely intending to refer to TR 281] (Davis) [death sentence after second sentencing proceeding affirmed; personal restraint petition denied]

TR 216 [likely intending to refer to TR 312] (Gregory) [direct appeal pending]

While Professor Beckett has now specified which cases she included in the retained category, she fails to explain the significance of this category or the associated percentages.

Summary: Dr. Scurich is critical of which denominator is listed in which column in the Updated Report, but this criticism goes to the clarity of the presentation and not to the substance. The numbers that are appropriate denominators in the "death penalty imposed" column of Table 2 depend on the information the table purports to present. The careful reader would understand, and Professor Beckett has clarified, that the denominators represent numbers and percentages related to all aggravated murder cases and not just those in which the State filed a death penalty notice.

Professor Beckett has adjusted her calculations of the percentage of special sentencing proceedings in which a death sentence was imposed by race of defendant, calculating that juries or judges imposed death in 38.8% of the cases involving death eligible nonblack defendants, but 64.3% of otherwise similar cases involving black defendants. If Trial Reports 92, 167, 182 and 224 should be excluded from these percentages, juries or judges imposed death in 42.6% of cases involving death eligible white defendants and 33.3% of cases involving death eligible "other race" defendants (combined, 41.3% of cases involving death eligible "nonblack" defendants), and 64.3% of cases involving death eligible black defendants.⁵⁰

Neither the ordinary meaning of the term "retained" nor the definition in the Updated Report and the Response to Evaluation provided a basis to verify the numbers in the "Death Penalty Retained" column of Table 2. Professor Beckett has now explained that the "retained" category encompasses cases in which a death sentence has been imposed and has not been reversed, including cases where a direct appeal has not been decided. She does not explain, and I do not discern, the significance of these percentages.

C. <u>Table 3: Capital Case Outcomes Among Death-Eligible Washington State</u> <u>Aggravated Murder Defendants, December 1981 - May 2014, by Race of</u> <u>Defendant and Race of Victim</u>

A note to Table 3 indicates the "[f]igures include only black and white 'death eligible' defendants with one white or black victim." *Updated Report* at 22. Dr. Scurich evidently read this note as including defendants with one or more white victims or one or more black victims. For example, he points out that the dataset shows 54 white defendants with all white victims, whereas Table 3 reports there are 33 such cases. *Evaluation of the Updated Report* at 12. Applying his understanding of Table 3, Dr. Scurich was unable to replicate the figures. Professor Beckett responds that Dr. Scurich misinterpreted the table note. She reiterates the explanation in the *Updated Report* that "Table 3 compares outcomes for black and white defendants convicted of killing a single white victim versus a single black victim." *Response to Evaluation* at 19 (quoting *Updated Report* at 23). She states the purpose of limiting

⁵⁰ As discussed above, one issue is whether two sentencing proceedings for the same defendant should be included in the regression analysis. I note that for purposes of calculating basic percentages, Professor Beckett included both the first and second sentencing proceedings where a defendant was sentenced to death by one jury, the sentence was vacated, and a second jury also imposed a death sentence. The calculations I set forth do not make any changes relating to the inclusion of multiple sentencing proceedings and do not adjust for Professor Beckett's recent inclusion of Trial Report 210.

the numbers to single victims was "to consider whether the descriptive data provided preliminary evidence that race of victim in combination with the race of the defendant may be consequential" and included only single-victim cases "in order to informally 'control for' the number of victims." *Response to Evaluation* at 18.

Additionally, Dr. Scurich asserts that the germane percentages should be based on a denominator of cases in which a death notice was filed. *Evaluation of the Updated Report* at 13. His original calculations using this denominator, and including cases with a white defendant and a white victim, had associated percentages that were "drastically different than the percentages reported in Table 3 of the Beckett and Evans report." *Evaluation of the Updated Report* at 14. It is now clear that Dr. Scurich's inability to replicate the results in the "Death Notice Filed" and "Death Penalty Imposed" columns of Table 3 was due to use of values different from those Professor Beckett identified in the Table 3 note and the accompanying narrative, as further explained in the *Response to Evaluation*. In answer to Interrogatory No. 16, Dr. Scurich states that once the germane values were more fully described, he was able to replicate the values in the "Death Notice Filed" and "Death Notice Imposed" columns of Table 3. *Answers to the Commissioner's Interrogatories* at 4.

Dr. Scurich also indicated that he "was <u>not</u> able to verify the numbers in the 'death penalty retained' column of Table 3, since this variable does not appear in the data file or the codebook." *Evaluation of the Updated Report* at 14 (emphasis in original). Like Dr. Scurich, I was unable to ascertain what was included in this category from the information contained in the *Updated Report* and the *Response to Evaluation*. As discussed above, in her responses to interrogatories Professor Beckett has specified the cases that were included in the "retained" category, but she does not explain the significance of the category and the associated percentages.

Summary: Dr. Scurich's inability to replicate the results in the "Death Notice Filed" and "Death Penalty Imposed" columns of Table 3 was due to use of values different from those identified in the Table 3 note and the accompanying narrative. With a fuller description of the values, he was able to replicate the values in these columns of Table 3. As to the "Death Penalty Retained" column, attempts to understand the source of these numbers did not reveal any consistent definition of "retained." Professor Beckett now explains that the "retained" category includes any case in which a death sentence has been imposed and has not been reversed, including cases where a direct appeal has not been decided. She does not explain, and I do not discern, the significance of these percentages.

III. REGRESSION ANALYSIS

Professor Beckett observes that "[i]t is conceivable that the racial differences described above are a function of case characteristics rather than of race itself. For example, if cases involving black defendants have, on average, more aggravating circumstances or fewer mitigating circumstances than cases involving white defendants, this could explain why juries sentence black defendants to death more frequently than they do white defendants." *Updated Report* at 23. Professor Beckett determined to use regression models to evaluate whether race is a cause of the disparity reflected in the basic percentages. Dr. Scurich acknowledges that regression techniques may be used in this context "to unconfound the factors that affect death sentences." *Evaluation of the Updated Report* at 27. However, Professor Beckett and

Dr. Scurich disagree on the validity of the results of Professor Beckett's application of the logistic regression techniques and on how the results should be interpreted.⁵¹

Regression models are used by many social scientists to account for multiple explanatory variables in a way that isolates the unique impact of one variable-here, the race of the defendant-in order to investigate a cause-and-effect relationship with the outcome. The outcome is called the "dependent variable," because it depends on the causes, while the variables that may represent the causes are called "independent variables." See "Reference Guide on Statistics" in Reference Manual at 219. Described in broad terms, these regression models are used to evaluate the possible reasons for an association between an identified independent variable and an outcome, recognizing that the association may result from (1) random chance,⁵² (2)a causal effect, or (3) confounding factors that may instead be the cause of the outcome.⁵³ The term "random" in this context includes factors at work in the decision making system that are unique features of individual cases that cannot be accounted for in any systematic way. See McCleskey v. Zant, 580 F. Supp. 338, 362 (N.D. Ga. 1984), aff'd in part, rev'd in part sub nom. McCleskey v. Kemp, 753 F.2d 877 (11th Cir. 1985), aff'd, 481 U.S. 279, 107 S. Ct. 1756, 95 L. Ed. 2d 262 (1987) (quoting expert testimony that "[t]he world really isn't random. When we say something is

⁵¹ Professor Beckett also studied whether the race of the defendant or race of the victim affected prosecutors' decisions to file death notices in eligible aggravated first degree murder cases. While critical of some aspects of this study, Dr. Scurich does not challenge Professor Beckett's conclusion that neither the race of the defendant nor race of the victim influenced prosecutors' decisions to file death notices. Accordingly, I do not address criticisms that are unique to these conclusions.

⁵² The impact of random chance on study results is also referred to as "random error." *See* "Reference Guide on Statistics" in *Reference Manual* at 240.

⁵³ The "Reference Guide on Statistics" in *Reference Manual* at 219 n.17 explains that a confounding variable may be correlated with an independent variable under study, and if the confounding variable acts causally on the dependent variable, the confounding variable (rather than the independent variable under study) could be responsible for observed effects on the dependent variable.

random, we simply mean it's unaccountable, and that whatever does account for it is unique to each case.").

Dr. Scurich does not challenge the appropriateness of the use of regression analysis to investigate the cause-and-effect relationship of race and imposition of the death penalty. However, Dr. Scurich asserts that the number of cases studied is fewer than the minimum sample size required for a regression analysis, and he challenges particular aspects of Professor Beckett's regression models and interpretation of the results. As discussed below, the relatively small dataset of cases is an unavoidable study feature that is below the minimum 100 observations considered necessary by the leading authorities on the maximum likelihood estimates (MLE) method used by Professor Beckett. However, courts generally consider issues of sample size to go to the weight accorded to a study rather than to its admissibility as evidence. Decisions on the probative value of the information developed are for the Justices, and it is important to set forth each issue that I considered in a manner that allows each Justice to independently decide the matter.

A. Probability Values (P-Values)

"A study that is statistically significant has results that are unlikely to be the result of random error. . . ." "Reference Guide on Epidemiology" in *Reference Manual* at 573. The first step in determining if the results are statistically significant is to select an appropriate p-value to use to evaluate the results of the particular study. Calculation of a p-value is a common method to evaluate the probability that chance affected the data such that the observed association was the result of random chance rather than a true association. *See* "Reference Guide on Statistics" in *Reference Manual* at 250.

Large p-values indicate that a disparity can easily be explained by the play of chance. If the p-value is very small, the data is so far away from the values expected under the "null hypothesis" (the hypothesis that there is no cause-and-effect) that statisticians would conclude that something other than chance must be involved. Stated another way, a small p-value indicates something is going on besides random chance. More specifically, a .05 p-value means that the probability is 5% of observing an association at least as large as that found in the study when in truth there is no association. *See* "Reference Guide on Epidemiology" in *Reference Manual* at 577. A .10 p-value means that the probability is 10% of observing an association at least as large as that found in the study at some association at least as large as that four p-value means that the probability is 10% of observing an association at least as large as that four p-value means that the probability is 10% of observing an association at least as large as that found in the study when in truth there is no association at least as large as that found in the study when in the study w

Dr. Scurich asserts that a p-value below .05 is generally considered necessary for a study result to be statistically significant.⁵⁵ See Evaluation of the Updated Report at 22 n.17. This .05 level is the most common p-value used for statistical significance in social science. "Reference Guide on Statistics" in *Reference Manual* at 251. Professor Beckett acknowledges this convention, but sets forth reasons why in the context of the analyses in the Updated Report she determined the p-values of the covariates as to sentencing decisions are statistically significant at both the .05 and .10 levels. These reasons relate to the interpretation of p-values and to the

⁵⁴ A caution is appropriate at this point: the p-value is not a basis to calculate the probability of the null hypothesis. Statisticians caution that under the frequency theory of statistics, there is no meaningful way to assign a numerical probability to the null hypothesis. In other words, a p-value that indicates a 5% probability of observing an association at least as large as that found in the study when in truth there is no association does not mean that there is a 95% probability that there is in fact an association. *See* "Reference Guide on Statistics" in *Reference Manual* at 252.

⁵⁵ Dr. Scurich indicates that he calculated both the p-values and 95% confidence intervals, and that if confidence intervals include the value of 1, the variable is interpreted as not being significantly predictive of the dependent variable. *Evaluation of the Updated Report* at 23. Professor Beckett responds that reporting p-values rather than confidence intervals is standard practice. *Response to Evaluation* at 6. Professor Beckett nevertheless adds 90% confidence intervals to her reports. *Id.* at 7. She reports the confidence intervals in a different manner than Dr. Scurich, such that in her computations the variable coefficient is not statistically significant if the confidence interval includes the value of 0. *Response to Evaluation* at 57. *Cf.* Dr. Scurich's *Answers to the Commissioner's Interrogatories* at 14-15 (explaining the dissimilar forms of reporting confidence intervals). Here, use of confidence intervals rather than p-values makes little difference in the experts' assessments of significance. In this report, I discuss p-values rather than the confidence intervals.

distinction between a one-tailed and a two-tailed test, as discussed in the following subsections.

1. Interpretation of Statistical Significance vs. Bright-Line P-Value

The experts disagree on whether, if the .05 level is selected, a p-value just above .05 nevertheless should be considered statistically significant, particularly in light of the American Statistical Association's recent "Statement on Statistical Significance and P-Values." This March 2016 statement sets forth principles underlying the proper use and interpretation of the p-value. One of these principles, quoted in Dr. Scurich's *Evaluation of the Updated Report* at 22, cautions that scientific conclusions and policy decisions should not be based only on whether a pvalue passes a specific threshold and that "[p]ractices that reduce data analysis or scientific claims or conclusions can lead to erroneous beliefs and poor decision making." The Statement explains: "A conclusion does not immediately become 'true' on one side of the divide and 'false' on the other. Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis."⁵⁶

As will be discussed below, Dr. Scurich maintains that the properly calculated p-value for the effect of the race of the defendant is above .05, and that as a result the finding that black defendants are more likely to receive a death sentence than nonblack defendants "disappears." *Evaluation of the Updated Report* at 4. Dr. Scurich's statement that a p-value below .05 is statistically significant and a p-value above .05 means the detected effect "disappears" does not comport with the principles

⁵⁶ The "Statement on Statistical Significance and P-Values" is available at http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108 (last visited November 20, 2017).

in the American Statistical Association (ASA) "Statement on Statistical Significance and P-Values," a statement which he quotes with approval.

Summary: Under the approach endorsed in the "Statement on Statistical Significance and P-Values," it is good practice to calculate the p-value and present this calculation together with other relevant factors to determine the probative value of the analyses. A p-value greater than .05 (or .10 for a one-tailed test) does not mean that the detected effect for the race of the defendant "disappears." Rather, the specific p-value should be considered with other study features that impact the degree to which study findings could reflect random chance rather than a true causal effect.

2. Choice of One-Tailed or Two-Tailed Test

The upper limit on a meaningful p-value may depend on whether it is appropriate to use a one-tailed or a two-tailed test. The terms "one-tailed" and "twotailed" refer to the "tails" or ends of a bell-shape curve that represents in graph form a "random normal distribution." See Palmer v. Shultz, 815 F.2d 84, 93 (D.C. Cir. 1987) (discussing the issue in detail and citing Wayne C. Curtis, Statistical Concepts for Attorneys at 72-73 (1983)). Professor Beckett explains these concepts in the Response to Evaluation at 43-45, and uses graphs to show the differences in one-tailed and twotailed tests. She explains that if the critical region for rejecting the null hypothesis lies at both ends of the distribution (a two-tailed test), the cut-off points for the .05 level of significance place 2.5 percent of the total area under the curve at the left end of the distribution and 2.5 percent at the right end of the distribution. Id. at 44-45. But "[i]f the critical region is concentrated at one end of the distribution, as it is in a one-tailed test, then all 5 percent of that area lies in one tail of the distribution." Id. at 44. An overall p-value of .10 is the equivalent of a p-value of .05 in one-tail of the normal probability distribution. To state the concept another way, if the hypothesis to be tested is one-tailed, such that all of the allowable probability for random chance is in one tail of the test, the statistically significant p-value may be twice as large as the statistically significant p-value for a two-tailed test.⁵⁷

Statisticians are wary of a study that sets the p-value at .10. See "Reference Guide on Statistics" in Reference Manual at 255-56 n.110 ("One-tailed tests at the 5% level are viewed as weak evidence-no weaker standard is commonly used in the technical literature."). Whether to use a one-tailed or a two-tailed test when testing for statistical significance is a common point of contention in cases related to statistical proof of discrimination. See Smith v. City of Boston, 144 F. Supp. 3d 177, 196-98 (D. Mass. 2015) reconsideration denied __ F. Supp. 3d_, 2017 WL 3184172 (2017) (discussing federal employment discrimination cases). In most circumstances experts and courts apply a two-tailed test, which is the more conservative approach. But as discussed below, statisticians recognize that there are circumstances where a onetailed hypothesis warrants consideration of a p-value of .10 or lower, with the admonition that the actual p-value should be considered together with other information in considering whether the null hypothesis can be rejected. See "Reference Guide on Multiple Regression" in Reference Manual at 321. Also, the selection of a one-tailed or two-tailed test may be more important if there is a brightline test of statistical significance, such that a p-value above a selected level cannot be considered statistically significant. If a bright-line test is not applied, the choice of a one-tailed or two-tailed test is not a critical threshold question. The "Reference Guide on Statistics" in the *Reference Manual* explains:

Some courts and commentators have argued for one or the other type of test, but a rigid rule is not required if significance levels are used as guidelines rather than as mechanical rules for statistical proof. One-tailed tests often make it easier to reach a threshold such as 5%, at least in terms of appearance. However, if we recognize that 5% is not a magic

⁵⁷ As Professor Beckett notes, in some circumstances researchers may present the results as .05 from a one-tailed test, but this indicates that the p-value was .10 and halved. *Response to Evaluation* at 44. Professor Beckett simply presents these results as a p-value of .10 to avoid an added layer of explanation and possible confusion.

line, then the choice between one tail and two is less important—as long as the choice and its effect on the p-value are made explicit.

Id. at 255 (footnote omitted). It is important that the p-value be made explicit because even if one does not employ a bright-line test for statistical significance, the degree to which the p-value is higher than .05 still bears on the overall question: whether there is an unacceptable level of probability that the study results indicating a disparity are the result of chance rather than a true effect.

A one-tailed test can be appropriate when there is a strong indication from evidence external to the statistical analysis that there is little chance that the data would point in the opposite direction from the one posited. As explained in the "Reference Guide on Multiple Regression" in the Reference Manual at 321, "By using a one-tailed test, the expert is in effect stating that prior to looking at the data it would be very surprising if the data pointed in the direct opposite to the one posited by the expert." Here, Professor Beckett states that her use of a one-tailed test analysis is premised on the view that it would be very surprising if the data pointed in the direction that black defendants were treated statistically better than nonblack defendants by sentencing juries, rather than the same or statistically worse than nonblack defendants. Cf. Palmer, 815 F.2d at 95 (placement of probability of randomness on the bell curve graph may depend on whether the possibility of discrimination that favors the protected group can be ruled out; if so, it can be said that the range of random disparities should be entirely within one of the tails of the bell curve rather than divided, with half of the range in each tail of the bell curve). If the hypothesis being tested were that being black neither increases nor reduces a defendant's likelihood of being sentenced to death, a two-tailed test would be appropriate. But if one can rule out the possibility of discrimination that favors black defendants and increases black defendants' likelihood of receiving a life sentence instead of a death sentence, the use of a one-tailed test would be appropriate.

Accordingly, Professor Beckett's use of a one-tailed test may be appropriate if evidence external to the statistical analysis strongly suggests that there is little chance that black defendants are treated less harshly, as opposed to equally or more harshly, than nonblack defendants. On a national basis, the evidence indicates that the criminal justice system contends with a "powerful racial stereotype—that of black men as 'violence prone.'" *See Buck v. Davis*, _____ U.S. _____, 137 S. Ct. 759, 776, 197 L. Ed. 2d 1 (2017) (quoting *Turner v. Murray*, 476 U.S. 28, 35, 106 S. Ct. 1683, 90 L. Ed. 2d 27 (1986) (plurality opinion)).⁵⁸ One might hope that these stereotypes do not significantly affect sentencing in capital cases in Washington, which has a different history than some regions of the country. But if evidence external to the statistical analysis strongly suggests the expectation that either black defendants are treated the same as white defendants or they are treated more harshly by Washington juries, and it would be very surprising if the data showed that black defendants were treated more favorably than nonblack defendants, use of a one-tailed test would be appropriate.⁵⁹

Summary: If one accepts the proposition that there is a powerful racial stereotype of black men as violence prone that may have influenced the juries' sentencing decisions in these cases, and it is unlikely that black defendants in these cases would have been treated more favorably by juries, there is no reason to reject Professor Beckett's use of the one-tailed test. The Justices of this court are in the best

⁵⁸ In *Turner*, 476 U.S. at 35, the plurality wrote: "Fear of blacks, which could easily be stirred up by the violent facts of petitioner's crime, might incline a juror to favor the death penalty." (Footnote omitted.) The amicus brief of the Constitutional Accountability Center submitted in *Buck v. Davis* references articles and studies on conscious and unconscious tendencies to consider black men as violent. This brief may be found at <u>http://www.americanbar.org/content/dam/aba/publications/supreme_court_preview/briefs_2016_2017/15-8049_amicus_pet_constitutional_accountability_center.authcheckdam.pdf (last visited November 20, 2017).</u>

⁵⁹ Even accepting the one-tailed approach, it should be noted that a p-value lower than .10 and closer to .05 would be stronger evidence that the results reflect a "true effect" and not a random result. As discussed above, the p-value is considered together with other factors to determine whether the study provides a basis for concluding there is a true effect.

position to evaluate the accuracy of this proposition as applied to Washington's justice system. If this proposition is accepted, a p-value less than .10 could support statistical significance, even if a rigid statistical significance level is required. If a rigid statistical significance level is not required, then the choice between a one-tailed and two-tailed test has less effect. Rather, the actual p-value may be considered together with other factors that bear on the question of whether the study reflects an actual causal effect rather than the operation of chance.

3. State Expert Challenges to and Variations of Professor Beckett's Regression Analyses and Associated P-Values

The experts' calculations of odds ratios and disagreements about the associated p-values have evolved as coding errors have been identified and missing information has been inserted in the data file. This section describes this process of correction and calculation as reflected in the experts' reports and in their answers to my questions. It concludes with a table that summarizes the "black defendant" odds ratios and associated p-values that Professor Beckett reports following correction of coding errors and insertion of information for "missing" values in the data file. As explained above, insertion of information that had been deemed missing but was in fact available resulted in the inclusion of additional cases in the models. At my request, these models use different assumptions as to which cases are properly included.

In the *Updated Report* at 30-31, Professor Beckett reported the results of a regression analysis of 75 cases in Table 7, "Impact of Case Characteristics and Defendant Race on Capital Sentencing Outcomes in Death Eligible Cases, December 1981 – May 2014." This table indicated that black defendants were 4.5 times as

likely⁶⁰ as nonblack defendants to be sentenced to death, after controlling for other variables included in the model, and that the results were statistically significant at a p-value of .05. In his *Evaluation of the Updated Report*, Dr. Scurich indicated that when he removed the "redundant" cases and then reran the model reported in Table 7, the effect for the variable representing the race of the defendant has an associated p-value of .062. *Evaluation of the Updated Report* at 26. However, this was not a salient rerun of the model, since it did not correct the identified coding errors. In Dr. Scurich's *Answers to the Commissioner's Interrogatories* at 4-5 he indicated that this model variation was run before he identified the coding errors and before he knew whether the coding errors would be acknowledged. In light of the confirmed coding errors, this rerun of the model with a resulting p-value of .062 did not provide meaningful information.

Dr. Scurich also reran the model after correcting the coding errors and removing the first sentencing proceedings for Rupe, Davis, and Gregory from the analysis. He reported a p-value of .053 for the effect of the race of the defendant. *Evaluation of the Updated Report* at 27. But Professor Beckett observed that it appeared that Dr. Scurich did not log the number of prior convictions or mitigating circumstances. Professor Beckett logged these variables on the basis they showed clear signs of skew. I make no pretense that I understand the underlying principles relating to transforming skewed variables by logging them. However, Professor Beckett is correct that such logging is generally recognized as an appropriate practice in standards of sound statistical analysis. For example, the "Reference Guide on

⁶⁰ In the Updated Report and Response to Evaluation, Professor Beckett used the phrase "more likely" in describing odds ratios, but in her Response to Commissioner's Supplemental Interrogatories uses the phrase "as likely." "As likely" is the appropriate term, since an odds ratio of 1 means the odds are the same. For example, as Professor Beckett explains in Response to Commissioner's Supplemental Interrogatories at 6 n.4, "[a]n odds ratio of 4.57 means that Black defendants are 357% more likely than non-Black defendants to be sentenced to death."

Multiple Regression" in the *Reference Manual* at 336 n.71 states that "[t]he logarithmic representation is appropriate when Y increases exponentially as X increases—for each unit increase in X, the corresponding increase in Y becomes larger and larger."⁶¹ Dr. Scurich does not dispute that a logarithmic representation is an established and appropriate method used when variables show signs of skew and does not claim that logarithmic representation is inappropriate for these variables. Rather, Dr. Scurich criticizes the *Updated Report* for not providing sufficient indication that a logarithmic transformation of certain variables (as relevant here, the number of prior convictions and number of mitigating circumstances) was used in the regression models, noting that "it appears that Beckett and Evans used a logarithmic transformation of these variables but never disclosed this fact in the Report."⁶² *Evaluation of the Updated Report* at 20. The criticism as to the lack of clarity on this point may be justified. Professor Beckett acknowledges that the notation "In" is used to show a logarithmic transformation, and "[w]e did not consistently list '(In)' after these variables in the tables; this was an oversight." *Response to Evaluation* at 27.

In order to ascertain the effect transforming the variables would have on the results of Dr. Scurich's model variation, I asked both experts to rerun Dr. Scurich's second model variation (in which he corrected coding errors and removed the first sentencing proceedings), but using Professor Beckett's logarithmic transformation of

⁶¹ Professor Beckett cites Agresti & Finlay, *supra*, at 561, who explain that in some circumstances logarithmic transformation "linearizes the relationship" in a way that is more useful in the analysis. *See Response to Evaluation* at 26 n.52.

⁶² Similarly, as to Tables 4-6, Dr. Scurich indicated he was not able to replicate findings for prior convictions or mitigating circumstances. Professor Beckett, suspecting he was unable to replicate the results because he did not transform the variables, confirmed this conclusion by replicating his models without the transformations and reaching Dr. Scurich's results. *Response to Evaluation* at 71-72. I note that the *Response to Evaluation* at 72 mislabels the table reproduced from the *Evaluation of the Updated Report* as relating to Appendix A7 (relating to Table 7) when in fact it is from Appendix A6 (relating to Table 6).

these variables.⁶³ Professor Beckett, while disagreeing that this is an appropriate model, calculated the p-value at .075. Dr. Scurich, while indicating he was "[a]ssuming *arguendo*" that values of 0 are to be replaced by 0.001 prior to the logarithmic transformation as described by Professor Beckett, reported a p-value of .072 for the race of defendant effect.

Professor Beckett had previously reported that when the model in Table 7 of the *Updated Report* was rerun with the three coding errors corrected, the Trial Report 34a added, and the first and second sentencing proceedings for Davis, Rupe and Gregory included, the p-value for the effect of the race of the defendant was 0.040. *See Response to Evaluation* at 64. I asked Professor Beckett to report the results if the first sentencing proceedings for Davis, Rupe and Gregory were excluded and Trial Report 34a was added to the dataset.⁶⁴ (Since Trial Report 34a was not

⁶³ Dr. Scurich had previously indicated that he "re-ran the model that appears in Table 7 [of the Updated Report], except that I used a logarithmic transformation of prior convictions and number of mitigating circumstances" but was unable to approximately replicate some of the significant findings from Table 7, including the effect of black defendant. Evaluation of the Updated Report at 20. The case processing summary in Appendix A7ii showed that only 55 cases were included in the analysis. See Evaluation of the Updated Report at 57-58. Professor Beckett posited that the 22 cases in which the trial reports listed no prior convictions and/or no evidence of mitigating circumstances were inadvertently omitted from the analysis. Response to Evaluation at 29. She explained that one cannot take the natural log of zero, and any case with missing variable values is automatically dropped from the analysis unless a very small number (such as .001) is added before applying the logarithmic transformation. Id. In his Answers to the Commissioner's Interrogatories at 6, Dr. Scurich notes that the Updated Report did not specify that values of 0 were replaced by a constant, but indicates the difference in the number of cases included in the model could account for the nonreplication. He further observes that there was no affirmative statement that .001 was the value used. Id. Professor Beckett confirms that "we added 0.001 to these zeroes." Responses to Commissioner's Interrogatories to Parties' Experts at 22.

⁶⁴ I asked Professor Beckett to provide the results of analyses that included and excluded Trial Report 34a in the event any argument emerged that it should not be included. As discussed above, the New Jersey Supreme Court's systemic proportionality review project initially included only one case for each defendant for statistical purposes, even if the same defendant committed murders on two separate occasions. But the special master and technical experts later concluded that the involvement of the same offender did not raise significant concerns about the lack of independence of the cases where the victims and circumstances of the cases were different. There does not appear to be a basis to exclude Trial Report 34a, and accordingly I summarize only the analyses that include this trial report.

included in the data file or the model reported in Table 7 of the Updated Report, it could not be included in Dr. Scurich's calculations. As explained above, Professor Beckett first added this case to the model reported in Table D in the Response to Evaluation.) Professor Beckett reiterated her position that it is inappropriate to exclude the first sentencing proceedings for Rupe, Davis, and Gregory from the analysis, but reported that for the model that excluded these cases and included Trial Report 34a, the p-value for black defendants was 0.066. See Professor Beckett's Responses to Commissioner's Interrogatories to Parties' Experts at 14.

However, as discussed above, it was subsequently determined that there were cases that were not included in the analyses because they were coded as having missing values when in fact the trial reports contained the values. At my request Professor Beckett has reported the results of analyses that include the cases for which values could be determined. The results of these analyses are presented in tables contained in Professor Beckett's *Response to Commissioner's Supplemental Interrogatories*. Table D1 presents the results obtained using the most recently updated data file (resulting in the inclusion of Trial Reports 8, 15, 210, and 313, which were previously excluded because of missing values) and including in the model all the cases Professor Beckett contends should be included. The results of this model are summarized in the first column below. The other models vary the set of cases used in the analyses so that the Justices will have the results of models that use different assumptions about the cases that are properly included. Professor Beckett does not agree that these additional models are appropriate, but has reported the results pursuant to my request.

[See table on following page]

	· ··			
Model Variation	<i>Table D1</i> Response to Suppl. Interrogs. at 5.	<i>Table D2</i> Response to Suppl. Interrogs. at 8.	<i>Table D4</i> Response to Suppl. Interrogs. at 11.	<i>Table D5</i> Response to Suppl. Interrogs. at 12.
Special Sentencing Proceedings Comprising Dataset for Analyses	81 special sentencing proceedings: <u>Includes</u> both original sentencing and resentencing proceedings for three defendants who were resentenced, and <u>Includes</u> the four special sentencing proceedings before judges where the State stipulated there were sufficient mitigating circumstances or that it could not meet its burden to prove the absence of sufficient mitigating circumstances	/8 special sentencing proceedings: Excludes original sentencing proceedings for three defendants who were resentenced (includes only their second sentencing proceedings), and Includes the four special sentencing proceedings before judges where the State stipulated there were sufficient mitigating circumstances or that it could not meet its burden to prove the absence of sufficient mitigating circumstances	77 special sentencing proceedings: <u>Includes</u> both original sentencing and resentencing proceedings for three defendants who were resentenced; and <u>Excludes</u> the four special sentencing proceedings before judges where the State stipulated there were sufficient mitigating circumstances or that it could not meet its burden to prove the absence of sufficient mitigating circumstances	74 special sentencing proceedings: Excludes original sentencing proceedings for three defendants who were resentenced (includes only their second sentencing proceedings), and Excludes the four special sentencing proceedings), and Excludes the four special sentencing proceedings before judges where the State stipulated there were sufficient mitigating circumstances or that it could not meet its burden to prove the absence of sufficient mitigating circumstances
Black Defendant Odds Ratio	4.568	4.001	4.072	3.558
P-Value	0.048	0.076	0.074	0.111

Summary: The experts disagree about which cases are appropriately included in the model reported in Table 7 of the Updated Report. I address the issues on which the experts disagree in other sections. Here I summarize the odds ratios and p-values that Professor Beckett reported for the race of defendant effect using variations of the model reported in Table 7 of the Updated Report and Table D of the Response to Evaluation. In these models the coding errors that the experts identified are corrected and missing values for which information is available have been inserted in the data file. Table D1 reflects the results of the model for cases that Professor Beckett maintains should be included. The other tables report the results of variations of the model to provide the Justices with information about the results of regression models that use different assumptions about what cases are properly included in an analysis of sentencing outcomes.

B. Small Dataset

Relatively few cases have advanced to special sentencing proceedings and fewer than half of those cases have resulted in imposition of the death penalty. In the *Updated Report*, Professor Beckett presented her results indicating that juries were more likely to impose a sentence of death when the defendant was black than in cases involving similarly situated nonblack defendants, and concluded "[a]lthough these results are based on analysis of a relatively small sample, they nonetheless indicate that the race of the defendant has had a marked impact on sentencing in aggravated murder cases in Washington State since the adoption of the existing statutory framework." *Updated Report* at 33. Dr. Scurich opined that the small "sample size" of cases in which the death penalty was sought and sentencing proceedings were conducted falls below the absolute minimum number needed to conduct an adequately powered statistical study. *Evaluation of the Updated Report* at 90-92. He discussed the concept that a study with a low statistical power has a reduced chance of detecting

a true effect, and cited an authority⁶⁵ that concludes that low power also reduces the likelihood that a statistically significant result that finds an effect reflects a true effect. *Evaluation of the Updated Report* at 92 n.57. Professor Beckett replied that the concept of "statistical power" is inapplicable, since the power of a statistical study relates to whether the study is able to detect an effect that exists, and her study clearly found an effect. *Response to Evaluation* at 22 Professor Beckett further responded:

We concur that a small number of cases is not ideal for logistic regression when results are intended to be generalized to or draw inferences about other populations. *This concern does not apply to our analysis*: these data are not a sample taken from a larger pool of cases, but rather encompass the entire population under study.

(Emphasis in original). This statement is followed by a footnote citing Alan Agresti and Barbara Finlay, *Statistical Methods for the Social Sciences*, 5-7 (3d ed. 1997). The cited authority defines "population" and distinguishes it from a "sample" by noting that "[t]he *population* is the total set of subjects of interest in a study. A *sample* is the subset of the population on which the study collects data." *Id.* at 4 (emphasis in original). It observes that predictions about the characteristics of a population can be based on information in a sample from that population, but there is no need for a sample when data exists for an entire population. *Id.* at 6. While the authority notes that sample size may be related to the likely accuracy of a sample statistic that predicts the value of a population, it does not address the question of whether a small number of cases is a concern where a regression analysis includes the population as opposed to a sample.

In considering the import of the small dataset, I began with the method Professor Beckett used to estimate that black defendants subject to special sentencing

⁶⁵ Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson & Marcus R. Munafò, *Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience*, 14 NATURE REVS.: NEUROSCIENCE 365 (2013).

proceedings are more likely than similarly situated nonblack defendants to be sentenced to death. Professor Beckett indicated that "we fitted logistic regression models, each with an outcome of 0 or 1, using Maximum Likelihood Estimate (MLE) procedures to estimate the probability of receiving a death notice or death sentence given a number of covariates." *Updated Report* at 16-17. Professor Beckett advised that "[i]n general, MLE estimates should be interpreted with caution for samples with fewer than 100 cases," referencing J. Scott Long, and Jeremy Freese, *Regression Models for Categorical Dependent Variables Using Stata*, (2nd ed. 2006).⁶⁶ *Updated Report* at 17. But Professor Beckett did not explain the principles of the chosen MLE procedures or the manner in which sample size may impact the results. Accordingly, the following explanation is based on my independent reading.

A technical explanation of MLEs is found in a recent edition of the referenced authority, J. Scott Long and Jeremy Freese, *Regression Models for Categorical Dependent Variables Using Stata* (3d ed. 2014), as follows:

ML [maximum likelihood] estimates are the values of the parameters that have the greatest likelihood of generating the observed sample of data if the assumptions of the model are true. To obtain the ML estimates, a likelihood function calculates how likely it is that we would observe the set of outcome values we actually observed if a given set of parameter estimates were the true parameters.

Id. at 84. Or as explained by another authority, "[t]his estimate is the value of the parameter that is most consistent with the observed data, in the following sense: if the parameter equaled that number (*i.e.*, the value of the estimate), the observed data would have had greater chance of occurring than if the parameter equaled any other

⁶⁶ This authority quotes an earlier publication by one of the authors, and states that "[i]t is risky to use ML[E] with samples smaller than 100, while samples over 500 seem adequate. These values should be raised depending on characteristics of the model and the data. First, if there are many parameters, more observations are needed. . . . A rule of at least 10 observations per parameter seems reasonable. . . . This does not imply that a minimum of 100 is not needed if you have only two parameters." *Id.* at 65 (alterations in original) (quoting J. Scott Long, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES. (1997).

number." Agresti & Finlay, *supra*, at 124. Since the MLE of a parameter is the value that makes it most likely to get the observed data, the probable accuracy of the MLE of a parameter relies on the distribution of the observed data points.

As I understand it, MLE procedures are generally used for larger datasets for the reason that chance may account for the distribution of the bulk of the data in a small dataset, in contrast to a large dataset where it is unlikely the distribution of the bulk of the data is due to chance. The general relationship of the size of a population under study and the operation of chance is explained in Allen v. United States, 588 F. Supp. 247, 418 (D. Utah 1984), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987). The court in Allen noted that there is a less potential for random chance to operate in the distribution of observations in a large population than there is in a small population, because "[i]n a large population, random variations tend to cancel each other out, yielding an overall observed distribution that is far more useful in evaluating correlations, relationships and probabilities." The court noted that this concept "may be demonstrated through the simple tossing of a coin." Id at 417. The probability of an evenly weighted coin turning up heads is 1'out of 2, but actually tossing the coin only a few times permits chance to operate in a significant fashion. Id. A coin tossed 10 times may, as a random matter, turn up heads 7 times. Id. But while random results are not unusual in a small set of tosses, it is unusual for random results to be sustained over a large number of tosses. A fair coin tossed 100 times is more likely to produce close to 50% heads than a test in which the coin is tossed only 10 times.⁶⁷ A similar example was used in a case involving claims of discriminatory employment practices. In United States v. City of New York [Vulcon Soc'y], 637 F. Supp. 2d 77, 95 (E.D.N.Y. 2009) the court noted that it has been recognized that

⁶⁷ This is a simplified summary of the court's use of a coin flip to explain probability theory. The coin flip analogy is used as an example in a discussion of the risk of "false positives" in the "Reference Guide on Epidemiology" in the *Reference Manual* at 576; this discussion notes the district court decision in *Allen*.
"[1]arger sample sizes create a greater likelihood that random differences between individuals will even out among all groups, and a lower likelihood that significant differences between the performance of racial or ethnic groups will have resulted from chance." The court quoted briefing by the United States that used a coin flipping analogy to explain how a large sample size enhances the reliability of statistical testing by moving from a small number of results that may be greatly impacted by chance to a large number of results that are less impacted:

"Flipping a fair coin 10 times will not always result in exactly five heads and five tails; a result of six heads and four tails on ten flips would not indicate with a reasonable degree of certainty that the coin was not fair (*i.e.*, that the disparity was not likely due to chance variation). However, if one flipped a fair coin 1,000 times, one would expect that the number of heads and tails would be close to equal, and a result of 600 heads and 400 tails would allow one to conclude with a high degree of certainty that the coin was not fair (*i.e.*, that disparity between the rate at which heads came up and the rate at which tails came up was not likely do [sic] to chance variation)."

Id. (quoting United States' memorandum in support of summary judgment). Applying this "fair coin" concept to a MLE, if there were 10 tosses with 7 turning up heads, the MLE of the parameter would be the degree of bias in the makeup of the coin that results in the observed 7 out of 10 tosses. If two of the tosses turned up heads due to random chance, the MLE of the parameter would not reflect a true effect. This appears to be the reason large datasets are required for use of a logistic regression model that employs MLE procedures to measure the causal effect of an independent variable on an outcome.

In the *Evaluation of the Updated Report*, Dr. Scurich did not discuss the MLE method. Rather, he discussed the issue of the small dataset in terms of "statistical power." Power is generally defined as the probability that a specified effect will be detected by the statistical hypothesis test, given that the effect exists. The "Reference Guide on Epidemiology" in the *Reference Manual* at 625 indicates that, in

less formal terms, power is like the strength of a magnifying lens in its capability to identify an association that truly exists. The basic principles of "statistical power" are set forth in the *Evaluation of the Updated Report* at 89-93. A study is underpowered if there is an insufficient dataset to reliably detect an effect. Power is a function of the sample size used in the study, the p-value threshold, and the effect size. Power equal to or greater than 80% is conventionally considered acceptable in the social sciences. A determination of the "sample size" needed is generally accomplished by selecting an 80% fixed level of power, a p-value threshold, a desired effect size, and then solving for the necessary sample size given the other selected assumptions. Dr. Scurich calculated that under the assumption the effect size is medium (*i.e.*, the odds ratio is 2.5), adequately powered regression analyses on the decision to impose the death penalty would require a sample size of 283. He calculates 138 cases would be required to detect a large effect (odds ratio of 4.0). *Evaluation of the Updated Report* at 90-91.

As this discussion indicates, the concept of statistical power is generally associated with the type of error that results in a "false negative." As both experts explain, there are two types of errors in null hypothesis statistical tests: (1) a Type I error, that inappropriately concludes the observed data supports rejection of the null hypothesis and the proposition that there is no difference in treatment (a false positive), and (2) a Type II error, that inappropriately concludes the data does not support rejecting the null hypothesis, thus failing to detect an effect that actually exists (a false negative). Since statistical power is generally defined as the probability that a test will appropriately reject the null hypothesis when the null hypothesis is false, thus avoiding a Type II error, it is difficult to conceive how the concept of statistical power applies to a study that has identified the presence of an effect. Professor Beckett points out this mismatch in concepts. *Response to Evaluation* at 50. However, an article cited by Dr. Scurich applies the concept of statistical power to "false positives" and the authors conclude that "[l]ow statistical power (because of low sample size of studies, small effects or both) negatively affects the likelihood that a nominally statistically significant finding actually reflects a true effect." Button, et al., *supra*, at 365-76 (2013). To a large extent, this conclusion appears to be based on a concept similar to the "coin toss" analogy discussed above in the context of the dataset needed for MLE procedures. The authors of the article focus on the fact that in a small study the influence of random chance makes it more likely one could find an inflated effect. *Id.* at 366-67 (discussing the "winner's curse" of a scientist who conducts a study of a small number of cases and who by chance gets results that inflate the true effect).

In response to Dr. Scurich, Professor Beckett discussed measures taken to address a small sample size, including identifying data points that are outliers and model testing. But nothing in the underlying reasons for these measures suggests that they change the minimum 100 observations for MLE procedures. Professor Beckett stated that "[w]hen conducting logistic regression analysis on a relatively small number of cases, it is important to ensure that neither outliers (*i.e.* highly unusual cases) nor small changes in model specification have undue influence on the results." *Response to Evaluation* at 22. An "outlier" is "[a] data point that is more than some appropriate distance from a regression line that is estimated using all the other data points in the sample." "Reference Guide on Statistics" in *Reference Manual* at 354. Identifying an outlier is an appropriate step, but it does not address the concern that chance may account for the distribution of the bulk of the data in a small dataset.⁶⁸ Professor Beckett's answer to my supplemental question on this point confirms that identifying and removing a data point that is an outlier does not address the extent to which chance or random variation accounts for the distribution of the remaining data for the purposes of MLEs. *See* Professor Beckett's *Responses to Commissioner's Interrogatories to Parties' Experts* at 17.

The other practice that Professor Beckett identified as necessary with a small number of cases is "model testing." Professor Beckett stated, "When conducting logistic regression analysis on a relatively small number of cases, it is of utmost importance to guarantee that small changes in model specification do not have undue influence on the results. It is precisely for this reason that we conducted rigorous model testing (or what [Dr. Scurich] calls sensitivity analysis) to determine what, if any, minor changes might impact the race of defendant effect." *Response to Evaluation* at 51. As Professor Beckett notes, sensitivity testing analyzes data in different ways to ensure that small changes in model specifications do not have undue influence on the results. *See* "Reference Guide on Statistics" in *Reference Manual* at 296 (defining sensitivity analysis). *Cf. In re Processed Egg Products Antitrust Litig.*, 312 F.R.D. 171, 189 (E.D. Pa. 2015) ("A sensitivity analysis is '[t]he process of checking whether the estimated effects and statistical significance of key explanatory variables are sensitive to inclusion of other explanatory variables, functional form,

⁶⁸ Professor Beckett states that "[d]iagnostic tests revealed one potential outlier" and that "[r]emoving this case from the analysis had no meaningful impact on the results." *Response to Evaluation* at 22 n.46. Professor Beckett notes the results with and without this outlier: "Specifically, after removing this case, the coefficient for Black defendant was 1.51 (4.54 times more likely than non-Black defendants to be sentenced to death after controlling for case characteristics and defendant race) with a p-value of 0.049. When the case is included, the coefficient for Black defendant was 1.573 (4.819 times more likely than non-Black defendants to be sentenced to death after controlling for case characteristics and defendant race) with a p-value of 0.040. All results presented in this report include this case." *Response to Evaluation*, Appendix A at 51 n.91. This description illustrates that the outlier did not have a significant impact on the results, but does not provide information on the distribution of the other data points.

dropping of potentially out-lying observations, or different methods of estimation"") (quoting Jeffrey M. Wooldridge, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH 845 (4th ed. 2009)). A statistic that does not change much when data or assumptions are modified slightly is "robust." "Reference Guide on Statistics" in *Reference Manual* at 295. *Cf.* "Reference Guide on Multiple Regression" in *Reference Manual* at 322-27 (discussing matters that relate to robustness of a model). But model testing does not relate to the minimum size of the dataset required for the MLE procedure. While Professor Beckett is correct that it is important to evaluate the robustness of regression analyses by evaluating the extent to which model results are sensitive to changes, such testing does not address the degree to which a small dataset provides data points that may be the product of chance. In response to my supplemental questions, Professor Beckett confirms that robustness and chance or random variation are separate issues. *See* Professor Beckett's *Responses to Commissioner's Interrogatories to Parties' Experts* at 19.

Professor Beckett generally accepts the description of MLE procedures that I provided in background to Interrogatory No. 23. See Commissioner's Interrogatories to Parties' Experts at 25. See also Professor Beckett's Responses to Commissioner's Interrogatories to Parties' Experts at 16. This description included the observation that "MLE procedures are generally used for larger datasets for the reason that chance variation may account for the distribution of the bulk of the data in a small dataset, in contrast to a large dataset where it is unlikely the distribution of the bulk of the data is due to chance variation." *Id.* But Professor Beckett adds, "To clarify, all data sets, large or small, contain some systematic variation that can be explained and chance or random variation, which cannot." *Id.* She reiterates that sampling error cannot account for the distribution of the data in her study, as follows:

Random variation is different than sampling bias. If researchers draw a sample from the relevant population, they must address the possibility that the distribution of the data in the sample is not fully representative

of the population they seek to understand. In this study, however, we are not analyzing a sample of Washington State capital cases. Instead, the data include *all* Washington State aggravated murder proceedings that took place from 1981 to May of 2014 for which trial reports are available. The data thus consist of the population of relevant proceedings rather than a sample of that population; sampling bias cannot account for the distribution of our data.

Id. This observation does not address the impact of chance factors unrelated to sampling that occur.⁶⁹ *Cf.* David C. Baldus, Catherine M. Grosso, George Woodworth & Richard Newell, *Racial Discrimination in the Administration of the Death Penalty: The Experience of the United States Armed Forces (1984-2005)*, 101 J. CRIM. L. & CRIMINOLOGY 1227, 1260-61 (2011) (noting that some may argue that tests of statistical significance are not relevant when the sample includes the "universe" of cases, but noting "we also believe that highly discretionary decisionmaking in a military criminal justice system is subject to random and unpredictable impacts whose potential effects can be assessed with tests of statistical significance").⁷⁰ Chance that is unrelated to sampling error is also a basis for requiring a larger number of observations for MLE. To return to the coin toss analogy, if one tossed the coin only 10 times and observed that the coin landed on "heads" 8 of the 10 times, this would comprise the full set of observations. The MLE for the bias in the makeup of the coin would be that degree of bias likely to produce the observed fact of 8 heads in 10 tosses. But to the extent that this small number of observations limits confidence that

⁶⁹ I note that in a small dataset of all decisions the effect of individual case coding errors could affect the reliability of the results produced in the models by not accurately including the values of the whole population of decisions. *Cf. State v. Loftin*, 157 N.J. 253, 290, 724 A.2d 129 (1999) (observing that "the relatively small size of the proportionality review database suggests that the cumulative effect of individual case errors could undermine the reliability of the results produced by the statistical models").

⁷⁰ However, Baldus, Grosso, Woodworth & Newell reject the view that disparities that do not meet a .05 or .10 level of statistical significance are irrelevant, stating, "In studies such as this, involving the universe of cases during a prescribed period of time, causal inference should be based on both the magnitude of the estimated effects and their statistical significance." *Id.* at 1263. Professor Beckett sets forth a similar view in her *Response to Commissioner's Supplemental Interrogatories* at 15; stating that where the dataset includes the universe of cases "the direction and size of the coefficients and magnitude of the odds ratios are most important; p-values are far less meaningful."

the calculated bias and the true bias are the same, the MLE is not considered an appropriate process to obtain a statistically significant estimate of the true bias. Cf. Baldus, Grosso, Woodworth & Newell, supra, at 1260-61 ("Although our sample includes the entire universe of death-eligible cases prosecuted in the Armed Forces during the period of our study, the total sample (n = 97) is small compared to similar studies of state court capital charging and sentencing. . . . Because our models fall on the low end of the guidelines accepted by statisticians, we place particular weight on our composite culpability scales, one of which-the salient factors scale-is completely independent of multiple regression results.") Cf. also Jon B. Gould & Kenneth Sebastian Leon, A Culture That Is Hard to Defend: Extralegal Factors in Federal Death Penalty Cases, 107 J. CRIM. L. & CRIMINOLOGY 643, 661 and 686 n.146 (2017) (explaining that researchers did not use the MLE method in study of relationships of independent variables and low defense resources where database included all 62 federal death penalty cases tried from 1998 to 2004 in which cost data were available because there were fewer than 100 cases, noting as follows: "The conventional approach to estimating regression models with a binary dependent variable is the multivariate logit model. However, logit models rely on maximumlikelihood estimation ('MLE'). MLE is based on large-sample theory, and it often performs poorly with small samples.").

Summary: The number of cases in the dataset of cases that advanced to special sentencing proceedings is fewer than the minimum number of cases (100) ordinarily required for application of the MLE procedures that Professor Beckett uses in analyzing sentencing outcomes. The reason a minimum number of 100 is considered acceptable is that chance may account for the distribution of the bulk of the data in a small dataset, in contrast to a large dataset where it is unlikely the distribution is due to chance. The distribution of the data is the basis for the measurement of the likely value of a parameter, since the MLE procedure estimates the values of a set of parameters under which the observed data would have the greatest chance of occurring. Accordingly, if the observed data occurred by chance, the MLE values of the parameters are in turn based on chance. In sum, the number of cases in the dataset for jury sentencing proceedings from December 9, 1981 - May 31, 2014, ordinarily would be considered too small to provide a basis for a statistically significant study. Professor Beckett addresses other concerns related to a small dataset, but her explanations do not demonstrate that the ordinary requirement for a minimum of 100 cases for MLE procedures is inapplicable. Even where a small dataset includes the whole population of decisions under study, the number of cases in the dataset may fall below what is necessary to provide a basis for a statistically significant study. The question of what weight or probative value to accord to the study in light of the small database and the other evidence presented is for the Justices to decide.⁷¹

C. Disclosure of Analyses Conducted

Based on his review of the Updated Report, Dr. Scurich concludes that "numerous analyses were conducted but not included in the final analyses or the

⁷¹ Claims of deficient sample sizes are generally viewed as going to the weight of a study rather than the admissibility of a study under ER 702. See U.S. Information Systems, Inc. v. International Broth. of Elec. Workers Local Union No. 3, AFL-CIO, 313 F. Supp. 2d 213, 232 (S.D.N.Y. 2004) ("despite the potential effect the small sample size could have on the persuasiveness of [an expert's] conclusions, his testimony is not inadmissible solely based on sample size alone"). Some courts have chosen to consider statistical studies of small samples or populations in conjunction with other evidence, even if the court determines the statistical evidence does not by itself establish a proposition, deciding the probative value and weight the study should be accorded in light of the statistical principles involved and the other evidence presented. See, e.g., Boston Chapter, N.A.A.C.P., Inc. v. Beecher, 504 F.2d 1017, 1020-21 (1st Cir. 1974). Cf. Allen, 588 F. Supp. at 418 ("That data from small populations must be handled with care does not mean that it cannot provide substantial evidence in aid of our effort to describe and understand events. Mathematical or statistical evidence, when properly combined with other varieties of evidence in the same case can 'supply a useful link in the process of proof.' If relied upon as a guide rather than as an answer, the statistical evidence offered in this case provides material assistance in evaluating the factual connection between nuclear fallout and plaintiffs' injuries.") (internal citation omitted).

Report, and no information was provided regarding how many analyses were actually conducted, the specific variables/configuration of the analyses, or any theoretical rationale for including or excluding variables other than they were not 'consistently relevant to the outcome.'" *Evaluation of the Updated Report* at 29 (quoting *Updated Report* at 18). Dr. Scurich emphasizes the need to disclose the number of hypotheses explored and all statistical analyses conducted in order for the reader to assess whether the hypotheses and analyses were legitimate tests for robustness, or were instead exploratory attempts to reach a desired result that is selectively reported. Dr. Scurich refers to exploratory analyses and selective reporting as "p-hacking." *Evaluation of the Updated Report* at 28.

In response, Professor Beckett presents the unaltered statistical output associated with the alternative models presented in the *Updated Report. Response to Evaluation* at 5 and Appendix C. This information is related to Professor Beckett's testing to determine if the results regarding the significance of the race of the defendant in decisions to impose a death sentence are robust across a variety of model specifications. *Response to Evaluation* at 5. Professor Beckett states that these alternative models were used for the legitimate purpose of testing the robustness of regression results.

But gaps remained in the explanation and information related to the selection of the predictor variables. In background to Interrogatory No. 26, I observed that Professor Beckett indicated that whether the victim was held hostage was one of the case characteristics selected for inclusion in the analysis of special sentencing proceeding decisions, *Updated Report* at 18, and that "[i]n these analyses, we treat evidence that the victim was held hostage as a measure of victim suffering," *Updated Report* at 18 n.60. Neither the *Updated Report* nor the *Response to Evaluation* explained why this treatment is appropriate. I observed that the *Codebook* contained a variable Judge_ProlongSuffInd for prolonged suffering as indicated by the trial

judge.⁷² This variable is distinct from the coding for "victim held hostage," which was separately coded. The Response to Evaluation indicated that sensitivity analyses were the basis for exclusion of the prolonged suffering variable, stating that "research assistants were asked to record whether judges indicated in words that a victim's suffering was prolonged or allowed to endure over time. However, this variable was not included in the final analyses because sensitivity analysis revealed that it was consistently insignificant." Response to Evaluation at 10 n.22. Accordingly, I observed the Response to Evaluation indicated that testing was conducted in relation to the selection of variables to include or exclude, at least as pertaining to the Judge_ProlongSuffInd variable, but this variable did not appear in any of the statistical output presented in the appendices to the Response to Evaluation. I requested that Professor Beckett provide a full description of the method and associated testing used in selecting which case characteristic variables to include or exclude in analyses of sentencing decisions and the results of such a selection method and testing. Professor Beckett provides this information in her Responses to Commissioner's Interrogatories to Parties' Experts at 19-44. As to the choice of the variable for victim suffering, Professor Beckett explains that ""[a]lthough we did test Prolonged Suffering in the model including only case characteristics, we chose to include Victim Held Hostage as our measure of victim suffering in the models we presented because we have greater confidence in this measure." Id. at 34-35. She discusses the difficulty experienced in capturing the compound nature of the trial report form question in a coding protocol. Id. at 34. Professor Beckett nonetheless provided regression results for a model in which this measure of Prolonged Suffering was included in place of Victim Held Hostage. Id. at 46. She reported that the odds

⁷² Question 4(h) on the trial report form asks the judge to "describe the nature and extent of any physical harm or torture inflicted upon the victim prior to death." This court has said that a brutal murder involving substantial conscious suffering of the victim makes the murder more deserving of the death penalty. *See State v. Cross*, 156 Wn.2d 580, 632, 132 P.3d 80 (2006) and cases cited therein.

ratio with use of Prolonged Suffering indicated that black defendants were more than four times as likely as similarly situated nonblack defendants to be sentenced to death, with an associated p-value of 0.064.

Summary: Dr. Scurich emphasized the need to disclose the number of hypotheses explored and all statistical analyses conducted in order for the reader to assess whether the hypotheses and analyses were legitimate tests for robustness, or were instead exploratory attempts to reach a desired result that is selectively reported. In response, Professor Beckett presented unaltered statistical output of all the models presented in the report as generated from the statistical software program used. I credit Professor Beckett's statement that these alternative models were used for the legitimate purpose of testing the robustness of regression results.

Additionally, Dr. Scurich observed that the models reported in the Updated Report were not the only models tested, and that information had not been provided regarding the other models tested and how variables were selected to include or exclude in the reported models. Professor Beckett has now provided a fuller description of the methods and associated testing used in selecting which case characteristic variables to include or exclude in analyses of sentencing decisions and the results of such selection methods and testing.

D. Use of Parsimonious Models

Due to the small size of the dataset of sentencing proceedings (as contrasted with the larger dataset of prosecutorial filing decisions), Professor Beckett developed "parsimonious models" to measure the possible role of race in sentencing. Professor Beckett did not use this phrase in the *Updated Report*, but introduced the underlying concept, noting as follows:

In the analysis of jury decision-making, we included case characteristics that would likely have been known by judges and jurors. These include:

the number of prior convictions possessed by the defendant; whether there were multiple victims; the nature of the defendant's plea (guilty vs. not guilty); the number of aggravating circumstances found by the judge or jury; the number of mitigating circumstances identified; the number of defenses offered; and whether the victim was held hostage. We also tested the significance of a number of social factors. Unfortunately, not all of these factors could be included simultaneously in the analysis of jury decision-making because the smaller sample size reduces the number of variables that can be included in the models. Model testing suggested that the only social factor that was consistently relevant to the outcome is the race of the defendant. For this reason, defendant race is the only social factor included in the analysis of sentencing decisions models presented here.

Updated Report at 17-18 (footnotes omitted). Additionally, Professor Beckett indicated that although the case characteristic of whether the defendant pleaded guilty was included in the model shown in Table 6, the model shown In Table 7 excluded this variable. Updated Report at 30-31. Professor Beckett explained, "In order to accommodate the addition of defendant-race, we did not include the nature of the defendants' plea in this model." Id at 30. Dr. Scurich indicates that he finds this omission of the "pled guilty" variable in the second model perplexing. Evaluation of the Updated Report at 89. The Response to Evaluation provides more detailed information about the endeavor to find "the most parsimonious model (the model with the fewest variables) possible that also included (or controlled for) all relevant case characteristics." Response to Evaluation at 21. More specifically, Professor Beckett notes the "rule of 10" for parsimonious models, and explains that this rule limits the number of covariates to a ratio of one independent variable per 10 of the least most frequent events. Response to Evaluation at 52. Professor Beckett states: "If we followed the most conservative approach to model building, we would limit the regression model of sentencing decisions (which resulted in 35 death sentences) to three independent variables (10/35 = 3.5)." Id.

Professor Beckett's explanation is consistent with the general approach other authorities have employed where a dataset is small.⁷³ However, I noted that Professor Beckett did not explain the logic or methodology for deciding that the "pled guilty" variable should be omitted and, relatedly, did not explain the basis for the conclusion that the model that omitted this variable and added the race of the defendant is a model that still "included (or controlled for) all relevant case characteristics." Response to Evaluation at 21. In Interrogatory No. 27 I asked Professor Beckett to provide a description of the method and any associated testing used in selecting which variable to remove from those included in the model shown in Table 6 in order to add the race of the defendant variable to the model shown in Table 7. In Interrogatory No. 28 I asked Professor Beckett to provide the basis for her statement that the model that omitted the "pled guilty" variable and added the race of the defendant is a model that still "included (or controlled for) all relevant case characteristics." Professor Beckett noted that entry of a guilty plea was among four case characteristics that did not show a significant relationship to sentencing outcomes during model testing conducted for the purpose of selecting variables to include in the model. See Professor Beckett's Responses to Commissioner's Interrogatories to Parties' Experts at 48. The other variables that did not show a significant relationship to sentencing outcomes were the following: number of prior convictions, whether there was one or more victims, and whether the victim was held hostage. Professor Beckett determined that one of these four variables should be removed in order to add the variable for defendant race. She elected to remove the "pled guilty" variable because "unlike the other case characteristics, a defendant's plea is not a pre-existing

⁷³ For example, the special master for the New Jersey Supreme Court's systemic proportionality review project, applied a "principle of parsimony" that "requires at least five, and more conservatively ten, of the less frequent outcomes per independent variable." *Baime Report III* at 17 (quoting Dr. John Tukey). *Cf. In re Proportionality Review Project (II)*, 165 N.J. 206, 216-18, 757 A.2d 168 (2000) (discussing the parsimonious regression model).

characteristic of either the crime or the defendant" and is "arguably unrelated to the question of culpability," and because of her understanding that the nature of the defendant's plea ("guilty" or "not guilty") cannot lawfully be the basis of the decision to impose a death sentence, noting this court's decision in *State v. Frampton*, 95 Wn.2d 469, 627 P.2d 922 (1981). *Responses to Commissioner's Interrogatories to Parties' Experts* at 48-49, 49 n.18.

Summary: Professor Beckett's use of a parsimonious model with a limited number of variables is consistent with the general approach other authorities have employed where a dataset is small. Dr. Scurich indicated that he was perplexed by the decision to omit the "pled guilty" variable in the second model, but did not challenge Professor Beckett's conclusion that the smaller dataset limits the number of variables that should be included in the models. Professor Beckett has set forth her reasoning for removing this variable in order to add to the model the race of defendant without increasing the number of variables. She explains that the choice was first narrowed to the four variables that did not show a significant relationship to sentencing outcomes. Among these variables, she believed the "pled guilty" variable was arguably unrelated to the nature of the defendant's culpability and understood that whether a conviction is based on plea of guilty or not guilty cannot lawfully be the basis of a decision on whether to impose the death penalty. Professor Beckett has now provided a full description of the methods and explanation of the reasons she chose to remove the "pled guilty" variable.

E. State Expert's Testing of the Sensitivity of the Race of Defendant Effect

Dr. Scurich tested the robustness of the reported effect that black defendants are more likely to receive the death penalty than nonblack defendants by running variants of the model Professor Beckett reported in the *Updated Report* at Table 7. Two of the four variants are discussed above, as these variants corrected

coding errors and removed "redundant" cases. Dr. Scurich ran two additional variants of Professor Beckett's model. One model variant categorized the race of the defendant into white, black, or other race, as opposed to considering black defendants versus nonblack defendants. Another model variant included a variable representing the race of the defendant and the race of the victim. I now consider these two model variants.

1. Separate Examination of White vs. Black and Other-Race vs. Black Defendants

Dr. Scurich ran a variation of Professor Beckett's model reported in the *Updated Report* at Table 7 in which he categorized the race of the defendant into white, black, or other, as opposed to black defendants versus nonblack defendants. *See Evaluation of the Updated Report* at 23. Based on the results of this model variant, Dr. Scurich concluded that "while Beckett and Evans purportedly detected an effect for black vs. all other defendants combined, it appears that black defendants are not more likely to receive a death sentence than white or other-race defendants individually." *Id.* at 24 (emphasis omitted). In response, Professor Beckett maintained that the use of black/nonblack categories was appropriate in light of the literature on the role of race in capital trials, and that in any event there were technical errors in Dr. Scurich's model variant. *Response to Evaluation* at 33. Specifically, Professor Beckett maintained the statistical output Dr. Scurich provided in his Appendix B1, *Evaluation of the Updated Report* at 61-64, showed that he failed to use logarithmic transformations of skewed variables, namely, prior convictions and mitigating

circumstances.⁷⁴ Dr. Scurich confirms that he did not use logarithmic transformations of the variables for prior convictions and mitigating circumstances. *Answers to the Commissioner's Interrogatories* at 8.

Dr. Scurich reran the model reported at 23 of the *Evaluation of the Updated Report* with both variables logarithmically transformed, the three identified coding errors corrected, and the first sentencing proceedings reported in Trial Reports 7, 180 and 216 removed. The values associated with a contrast of black defendants with white defendants are similar to the values Dr. Scurich reports when the model reported in Table 7 of the *Updated Report* is rerun with these same changes, but with black defendants compared to white defendants and other-race defendants combined. *Compare* output table at 6 of Dr. Scurich's *Answers to the Commissioner's Interrogatories* at 6 (reporting a p-value of .072 and odds ratio of 4.115 for black defendant compared to nonblack defendant) with output table at 11 (reporting a pvalue of .077 and odds ratio of 4.083 for black defendant compared to white defendant). Dr. Scurich concludes that the contrast with white defendants is not statistically significant for the same reason he concludes that the contrast with

⁷⁴ In relation to this model variant, Professor Beckett indicated that "one important correction must be made to Dr. Scurich's comment about interpreting confidence intervals for log odds coefficients." *Response to Evaluation* at 49. She emphasized that when log odds (as opposed to odds ratios) are reported, coefficients that are a positive number indicate a negative relationship, coefficients that are a negative number indicate a negative relationship, and a value at or approaching 0 is the value that indicates no association. *Id.* But Dr. Scurich indicated that he included confidence intervals around Exp(B) which he describes as "the exponentiation of the logarithmic (natural log) beta parameter" or "[i]n short, it is an odds ratio." *Evaluation of the Updated Report* at 23. Dr. Scurich referenced the confidence intervals for Exp(B) when he stated that confidence intervals that include the value of 1 indicate that the associated odds ratio could be 1:1 (neither increasing nor decreasing the likelihood of the dependent variable) and the variable is interpreted as not being "significantly" predictive of the dependent variable. *Id.; see also Answers to the Commissioner's Interrogatories* at 14-15.

combined white and other race defendants is not statistically significant—because he maintains that a p-value above 0.05 is not statistically significant.⁷⁵

Summary: When Professor Beckett's logarithmic transformations are used, there is little difference in the p-values and odds ratio for black defendants compared to white defendants and black defendants compared to nonblack defendants as a group.

2. Race of Victim

Dr. Scurich ran another model variant based on his premise that the race of the victim should have been included in the model. As a general matter, failure to include an explanatory variable that is correlated with the variable of interest in a regression model may cause an included variable to be credited with an effect that is actually caused by the excluded variable. "Reference Guide on Multiple Regression" in *Reference Manual* at 314. Here, the hypothesis could be that failure to include the race of the victim may cause the race of the defendant to be credited with the effect of imposition of the death penalty when such an effect is actually associated with the race of the victim. Professor Beckett indicated she assessed whether the race of the victim influenced prosecutorial and/or jury decision-making in capital cases adjudicated in Washington. She concluded that neither the race of the defendant nor the race of the victim appeared to affect prosecutorial decision-making in aggravated murder cases and the fact that a victim was white was not a significant factor in sentencing outcomes. *See Response to Evaluation* at 30.

⁷⁵ In her *Response to Evaluation* at 33-34 Professor Beckett offers an alternative way to compare black defendants with white defendants. Dr. Scurich questions the accuracy of the reported p-value and also argues the methodology is flawed. *Answers to the Commissioner's Interrogatories* at 9-10. In light of my conclusion that Dr. Scurich's model variant does not demonstrate that there is a significant difference in results of regression analyses when black defendants are compared to white defendants, this aspect of the challenge is unfounded and it is not necessary to address Dr. Scurich's criticisms of the alternative model Professor Beckett presented to respond to this challenge.

It is against this background that I consider Dr. Scurich's position that the race of the victim is a "theoretically relevant variable" that should have been included in the model, *Evaluation of the Updated Report* at 22, and his conclusion that "once the race of the victim is accounted for in the model, there are no racial effects – for either the victim or the defendant – with respect to the imposition of the death penalty," *id.* at 25 (emphasis omitted). Dr. Scurich concludes that "when the race of the victim as well as the race of the defendant is included in the model, neither the race of the victim nor the defendant is related to receiving a death sentence." *Id.* at 3. Dr. Scurich reaches this conclusion after creating a "DefRaceXVicRace" variable, which considers combinations of defendant race and victim race.

As I set forth in background to my interrogatories, I do not understand how a model using this variable can support a conclusion that detected effects for the race of the defendant in regression analyses are reduced once one considers the race of the victim, even where there is no detected effect for the race of the victim. The variable that considers combinations of defendant and victim race may be appropriate to investigate whether juries and judges are more likely to impose the death penalty because of race-based opprobrium when a defendant of one race kills a victim of another race, or to investigate whether there is a lesser degree of condemnation when a defendant of one race kills a victim of another race. For example, some studies using a "DefRaceXVicRace" variable might show that black defendants are more likely than white defendants to receive the death penalty when victims are white. Or a study using this variable might show that white defendants are more likely to receive the death penalty when the victim is white than when the victim is black. But consideration of race in assessing whether the defendant should receive the death penalty is antithetical to constitutional principles whether that consideration is of the race of the defendant alone or in combination with the race of the victim.⁷⁶ The race of the victim is thus a "theoretically relevant variable" only to identify whether an alternative illegitimate factor may have influenced the outcome of sentencing proceedings.

Accordingly, in background to interrogatories I indicated that I did not understand the logic of an analysis like Dr. Scurich's that begins with Professor Beckett's regression model that predicts black defendants are more likely to receive a death sentence, varies the model by including the race of the victim, and emerges with the conclusion that once the race of the victim is accounted for any racial effects disappear. This analysis seems incongruent with the basic rationale of using regression techniques in this context, which is to unconfound the factors that influence a jury's decision to impose a death sentence by controlling for other legitimate case characteristics and then considering the extent to which a defendant's odds of being sentenced to death are enhanced by virtue of illegitimate characteristics. Logically, one would want to know whether the race of the defendant continues to be a significant predictor of imposition of the death penalty after factoring in other case characteristics-such as whether the victim was held hostage or the lack of a significant criminal history-that could influence the outcome and explain the results based on constitutionally sound decision making. From a legal view, considering either the race of the defendant or the race of the victim equates to placing greater or lesser value on the individual according to his or her race. And I cannot conceive of circumstances where a jury could consider the combination of the race of the defendant and the race of the victim in a manner that the law would find acceptable. I outlined these concerns in background to Interrogatory No. 30 and asked Dr. Scurich

⁷⁶ In *McCleskey v. Kemp*, 481 U.S. 279, 308-09, 107 S. Ct. 1756, 1776, 95 L. Ed. 2d 262 (1987), it was taken as a given that prejudice based on the victim's race is not a factor that may legitimately influence decision in the criminal justice process; the question there was at what point the risk of racial prejudice influencing capital sentencing decisions became constitutionally unacceptable.

to explain the theoretical basis for the position that a model variant that accounts for the race of the defendant and the race of the victim demonstrates there is no racial effect for either the defendant or the victim when the model that accounted only for the race of the defendant showed an effect. He responds that "[t]he non-significance of 'DefRaceXVicRace' implies that defendant race as well as victim race is not related to death sentences'' and further states as follows:

Regarding the theoretical basis of the variable DefRaceXVicRace, note that Table 3 of Updated Report at 22 disaggregates the data by both race of defendant and race of victim concurrently, and the associated text refers to different "racial configurations" of defendant and victim race (*Updated Report* at 22). These "configurations" refer to an interaction in statistical parlance. This led me to infer that an interaction of the race of the defendant and the race of the victim should be accounted for in the model.

Dr. Scurich's Answers to the Commissioner's Interrogatories at 12. This response does not address my questions about the theoretical basis of the model variant as it relates to consideration of race in sentencing proceedings. I remain unconvinced that a regression model reflects actual decision processes when it predicts black defendants are more likely to receive a death sentence, varies the model by including the race of the victim, and emerges with the conclusion that once the race of the victim disappear. Perhaps there is a technical answer to why the results of models that include both individual defendant race classifications and defendant/victim racial combinations are illogical. *Cf.* Baldus, Woodworth, Zuckerman, Weiner & Grosso, *supra*, at 168-69 (discussing race variables that address concern about discrimination against racial minority defendants and against racial minority defendants whose victims are white, but cautioning that "models which mix main effect race variables and race-based interaction terms create risks of error in interpretation in the absence of a more complicated coding regime and interpretative strategy").

Summary: Dr. Scurich fails to explain how his variant of a model that shows effects for the race of the defendant results in a conclusion that there were no racial effects for either the race of the defendant or the race of the victim. He offers no explanation of how adding consideration of victim race suggests race neutral reasons for sentencing decisions.

F. Interpretation of R² and Pseudo R² Measures

In the Updated Report, Tables 4, 5 and 6 include an R-squared (R^2) measure, while Table 7 includes a "Pseudo R^2 " measure. An R^2 statistic is the percentage of variation in the dependent variable that is accounted for by all the explanatory variables used in the particular model. "Reference Guide on Multiple Regression" in *Reference Manual* at 345. R^2 varies between 0 (the explanatory variables explain none of the variation of the dependent variable) to an R^2 of 1 (the explanatory variables explain all of the variation of the dependent variable). *Id.*

Dr. Scurich questions the use of the R^2 statistic in the context of logistic regression, as follows:

The Beckett and Evans report does report " \mathbb{R}^2 " for each logistic regression model, which is described as the "proportion of variation in outcomes explained." This is an appropriate description of " \mathbb{R}^2 " (R-squared) for linear regression but not logistic regression. As others have noted, "numerous formulas have been devised to yield an equivalent of this concept for the logistic model. None, however, renders the meaning of variance explained. Furthermore, none corresponds to predictive efficiency and none can be tested in an inferential framework." Thus, statements by Beckett and Evans such as "adding social factors to the model more than doubles the proportion of variation in outcomes explained (to 20%)" are plainly incorrect.

Evaluation of the Updated Report at 85-86 (footnotes omitted; internal quotation is noted as from Peng, C. Y. J., & So, T. S. H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 1(1), 31-70 at 45).

Professor Beckett responds by showing "Pseudo R^{2} " statistics rather than R^{2} statistics in all model results presented in the *Response to Evaluation*, explaining the term "Pseudo R^{2} " as follows:

Pseudo R2 - This is the pseudo R-squared. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics. Because this statistic does not mean what R-square means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic only to compare models.

Response to Evaluation at 56 (emphasis added). Presumably this comparison is to determine which particular model provides the best "goodness of fit" by comparing the percentage of variation in the dependent variable that is accounted for by explanatory variables used in the models.

Several statements in the *Updated Report* indicated the R² statistic was used to assess the percentage of variation in the dependent variable that is accounted for by the case characteristics that are included in a particular model. Professor Beckett explained the general concept as follows:

For each set of regression analyses, we first report the results obtained when only case characteristics are included in the model. This allows us to identify which case characteristics influence decision-making in death-eligible cases; *it also allows us to assess the proportion of the variation in outcomes that is explained by case characteristics as a group*.

Updated Report at 18 (emphasis added). After reporting results of " $R^2 = 0.0914$ " for a model (Table 4) that included only case characteristics that Professor Beckett identified as influencing prosecutorial decisions to file death notices, she concluded:

These results show that the case characteristics included in the model explain a small proportion (just 9%) of the variation in whether prosecutors file a death notice. In other words, most of the variation in prosecutorial decisions regarding whether to seek the death penalty is *not* a function of the case characteristics included in this model.

Updated Report at 25 (emphasis in original). Professor Beckett then added social factors to the model (Table 5), obtaining results of " $R^2 = 0.2063$ " and concluded that "adding social factors to the model more than doubles the proportion of variation in outcomes explained (to 20%)." Id. at 27. She then summarized that, with certain exceptions, "[o]verall, these results indicate that case characteristics alone explain a very small proportion of the variation that characterizes prosecutorial decisions about whether to seek the death penalty." Id. at 28. Similarly, as to jury decision to impose a death sentence, Professor Beckett interpreted the reported " $R^2 = 0.2117$ " results of a model that included only selected case characteristics as "[i]ndicat[ing] that case characteristics explain 21% percent [sic] of the variation in decisions to impose the death penalty." Id. at 29. After adding defendant race to the model and reporting results of "Pseudo $R^2 = .2473$," Professor Beckett stated: "Adding data regarding defendant-race notably improves the model: the amount of variation explained increases from 21 to 25 percent." Id. at 30. Professor Beckett indicated that these results supported the second of her three main conclusions, which she presented as follows:

Second, the regression results indicate that case characteristics explain only a small proportion of the variation in the case outcomes analyzed here. Specifically, case characteristics alone explain only 9% of the variation in prosecutorial decisions regarding whether to seek death and 20% of the variation in juries' sentencing decisions. Four case characteristics were significant predictors of prosecutorial decisions to file death notices: the number of prior convictions possessed by the defendant, the number of aggravating circumstances alleged by prosecutors, evidence that the defendant was suspected of committing a sex crime in the course of the homicide, and the involvement of law enforcement officer victims. Neither the number of victims nor evidence that the victim was held hostage were found to be significant predictors of prosecutorial decisions to file a death notice. Several case characteristics were also significant predictors of the decision to impose a sentence of death: the number of applied aggravating circumstances, the number of mitigating circumstances, the number of defenses, and whether the victim was held hostage. Overall, however, the case characteristics for which data are available and which are presumed to be the primary drivers of decision-making in capital cases actually explain a small proportion of the variance in case outcomes in aggravated murder cases. Unexplained variation documented in the results presented here suggest that other extra-legal and social factors – not captured by our statistical models – are playing an important role in death penalty case dynamics.

Id. at 32 (emphasis added).

In the *Response to Evaluation*, Professor Beckett appeared to retreat from the position that the results reported as R^2 or "Pseudo R^2 " are properly interpreted as showing the proportion of variation in outcome explained by the variables included in the regression model. The *Response to Evaluation* at 56 reports all results as "Pseudo R^2 " and states, "Because this statistic does not mean what R-square means in OLS [ordinary, i.e., linear, least squares] regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic only to compare models." This statement appeared to undercut the basis for all statements in the *Updated Report* that the case characteristics included in the models explain only a small proportion of the variance in prosecutorial decisions and sentencing outcomes, and her suggestion that this indicates that extra-legal and social factors other than those accounted for by case characteristics are playing an important role in death penalty decisions. I asked Professor Beckett the following question as Interrogatory No. 32:

Do you maintain that the results of models presented in the *Updated Report* provide a basis to determine the percentage of the variance in outcome that is explained by the case characteristics included in the models? If yes, please identify the models and explain.

Professor Beckett responded:

No, we do not maintain that the models presented in the Updated Report provide a basis to determine the percentage of explained variation. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression. However, many people have tried to develop one, and there are a wide variety of pseudo-R-square statistics. Although Pseudo R-squared statistics cannot be interpreted independently or compared across datasets, they are valid and useful in evaluating multiple models predicting the same outcome using the same dataset. For these reasons, and because this statistic does not mean what R-square means in OLS regression (the proportion of variance explained by the predictors), we suggest using this statistic only to compare models using the same dataset. In this situation, the higher pseudo R-squared identifies the model that better predicts the outcome. Professor Beckett's *Responses to Commissioner's Interrogatories to Parties' Experts* at 51 (footnote citing authority omitted). Professor Beckett adds that "[a]lthough none of the Pseudo R-squared measures can be interpreted as an exact percentage of variation explained, none of these values approaches 1.0, indicating there is much unexplained variation in the decision to impose death." *Id.* at 54.

Summary: Professor Beckett has withdrawn her use of R^2 statistics to indicate the percentage of variation in prosecutorial decisions or sentencing decisions that are accounted for by the variables used in the regressions models. "Pseudo R^2 " statistics are used instead, and Professor Beckett recommends interpreting these statistics only to compare models using the same dataset.

IV. CONCLUSION

This report sets forth information that is intended to facilitate the Justices' evaluation of the parties' expert submittals, recognizing that any decisions regarding the significance of these submittals is for the Justices. Submission of this final report concludes the hearing on the State's challenge to the *Updated Report* that the Court ordered be held before me.⁷⁷

ACTING DEPUTY COMMISSIONER

November 21, 2017

⁷⁷ The parties' were previously advised that I retired from the position of Commissioner effective June 16, 2017; my appointment as an acting deputy commissioner has provided authority to complete this hearing and report. *See* SAR 15(l) (deputy commissioners shall have the power to perform any act or duty relating to the commissioner's office that the commissioner has the power to perform).