

**This form effective September 15, 2022**

**Existing Data** may be in the form of individual records (e.g., academic, medical, financial), data sets, interview notes, biospecimens, online profiles and posts (e.g., social media), and audio- or video-recordings. These data could be available for purposes other than research, and are sometimes, but not always, identifiable.

This form is to be used when:

1. The data are individually identifiable. Data are identifiable if they include direct identifiers, such as names, email addresses or images. They may also be identifiable if a person’s identity could be deduced based on the information contained in the data (e.g., a combination of unique demographics characteristics).

Note: Data are considered identifiable when they contain unique identifiers linked to participants’ names, even if the researcher does not have the key.

1. The data provider requires IRB approval even though the data are not identifiable.
2. The data, whether identifiable or not, are not public. Data are not public if researchers need to create an account, register a study, submit an application, or enter into an agreement before accessing the data. Data provided by colleagues by way of a personal arrangement or request are not considered public.

This form is to be used **regardless of the type of review** (exempt, expedited, or full) that will be used for the protocol.

* Fill out this application and email it to campusirb@duke.edu or the IRB staff person you are working with. **Combine this form and any appendices into a single Word file.** Submitting the protocol in multiple files or as a PDF will delay the pre-review of your application. (However, agreements specifically pertaining to the request or access of existing data can be submitted as PDFs.)
* **Please note**: Signatures in Section 2 are required for final approval. Faxed and scanned signatures are acceptable, as are electronic signatures. We cannot accept typed names.

|  |
| --- |
| Section 1: General Information |

**Protocol Title**: Understanding Prosecutorial Discretion

|  |
| --- |
| Section 2: Key Study Personnel |

**Principal Investigator**

Identify one Principal Investigator (PI) on this project and sign below.

* This person is responsible for the overall conduct of the research. For all students, fellows, and post-docs, this is your faculty advisor
* If you have more than one PI, only choose one
* By signing, the PI certifies to the following:
  + I have read and approved the protocol
  + I assume responsibility for ensuring that my advisees are aware of the responsibilities as researchers
  + I ensure that the IRB will be immediately notified in the event of [unanticipated risks to participants, protocol deviations, or findings during the study that would affect the risks](https://campusirb.duke.edu/node/101) of participation.

|  |  |
| --- | --- |
| **Name**: **Nicholas Eubank** | **Department or School**:  Duke Social Science Research Institute (SSRI) |
| **E-mail Address**: nce8@duke.edu  **NetID**: nce8 | **Phone Number**: 303-918-6528 |
| **Faculty Advisor**  **Faculty Researcher**  **Staff**  **Other**:Click or tap here to enter text. | |
| **Signature**: | **Date**: 5/17/2023 |

**Duke Research Team**

Please list the other Duke members of the research team AND indicate their role on the project. Do not list non-Duke researchers. These team members can be added in a later section.

Feel free to copy and paste, or delete the blocks as necessary.

All signatories agree to the following:

* I will not begin the research until written approval is secured from the IRB. Note: Approval will not be provided unless [certification to conduct research with human subjects](https://campusirb.duke.edu/node/57) for each researcher named on the protocol is current.
* I will conduct this study as described in the approved protocol.
* If any changes are anticipated, I will submit a [Request to Amend an Approved Protocol](https://campusirb.duke.edu/node/22), and I will not implement the changes until I receive approval from the IRB.
* I will contact the IRB staff promptly if any of the following events occur: [unanticipated risks of harm to participants, protocol deviations, and findings during the study that would affect the risks](https://campusirb.duke.edu/node/101) of participation.

|  |  |
| --- | --- |
| **Name**:Clarissa Ache Cabello | **Department or School**:  Social Science Research Institute |
| **E-mail Address**: clarissa.ache.cabello@duke.edu  **NetID**: ca244 | **Phone Number**: 3059278651 |
| **Faculty**  **Undergraduate**  **Graduate** **student**  **Postdoc**  **Research associate**  **Other**:Just graduated (May 2023) Duke Masters of Interdisciplinary Data Science student | |
| **Signature**: | **Date**: 5/18/2023 |

|  |  |
| --- | --- |
| **Name**:Click or tap here to enter text. | **Department or School**:  Click or tap here to enter text. |
| **E-mail Address**: Click or tap here to enter text.  **NetID**: Click or tap here to enter text. | **Phone Number**: Click or tap here to enter text. |
| **Faculty**  **Undergraduate**  **Graduate** **student**  **Postdoc**  **Research associate**  **Other**:Click or tap here to enter text. | |
| **Signature**: | **Date**: Click or tap to enter a date. |

|  |  |
| --- | --- |
| **Name**:Click or tap here to enter text. | **Department or School**:  Click or tap here to enter text. |
| **E-mail Address**: Click or tap here to enter text.  **NetID**: Click or tap here to enter text. | **Phone Number**: Click or tap here to enter text. |
| **Faculty**  **Undergraduate**  **Graduate** **student**  **Postdoc**  **Research associate**  **Other**:Click or tap here to enter text. | |
| **Signature**: | **Date**: Click or tap to enter a date. |

**If there are more members of the research team, copy and paste the researcher information and signature block as needed.**

**Other Study Contacts**

If there are additional personnel (e.g. a departmental staff member) who assist in protocol preparation and record keeping, and would like to be copied on correspondence from the IRB, please add them here.

|  |
| --- |
| **Name**:Click or tap here to enter text. |
| **E-mail Address**: Click or tap here to enter text.  **NetID**: Click or tap here to enter text. |
| **Type of Correspondence:**  Approval and Reminder Notices  All correspondence related to the submission, including feedback |

**IRB USE ONLY**

This section is to be completed by IRB staff or IRB members only.

|  |  |
| --- | --- |
| **APPROVED as**  **Exempt**  **Expedited or**  **Full** | |
|  |  |
| IRB Designee or  IRB Member | Date |

|  |
| --- |
| Section 3: Departmental & Institutional Affiliations |

1. **Identify the department, institute, or center that you consider the home of the study.**

|  |
| --- |
| Duke University Social Science Research Institute |

1. **Will you be collaborating with researcher(s) at other institution(s)?**

See our [**Collaborative Research**](https://campusirb.duke.edu/node/103) policy or contact IRB staff at [campusirb@duke.edu](mailto:campusirb@duke.edu) to determine whether you are engaged in an inter-institutional collaboration.

Yes  No

**If YES, please specify the following for each collaborator:**

|  |  |
| --- | --- |
| Collaborator’s Name: | Click or tap here to enter text. |
| Role in Research: | Click or tap here to enter text. |
| Research Activities/ Responsibilities: | Click or tap here to enter text. |
| Organization/Institution: | Click or tap here to enter text. |
| Has your collaborator reached out to their organization/institution about IRB or ethics review? | |
| Yes\*  No  Collaborator’s organization/institution does not have an IRB or ethics review board | |
| \*If you indicated that your collaborator has reached out to their organization/institution’s IRB or ethics review board, please describe their determination or the status of the request:  Click or tap here to enter text. | |

**Important items to note:**

* Review by the Campus IRB extends only to members of the Duke research team unless there is a formal, fully executed inter-institutional agreement among the IRBs engaged in the research.
* If your collaborator is a foreign entity, IRB staff will forward this protocol to the [Export Controls](https://export.duke.edu/) and [Privac](https://oarc.duke.edu/privacy)y offices for ancillary reviews.

|  |
| --- |
| Section 4: Funding Sources and Conflict of Interest |

1. **Please identify your funding source(s)**:

|  |
| --- |
| N/A |

1. **Are any of the above funding source(s) a U.S. Federal Agency or Department?**

Yes  No

**If YES, please append the grant application to this protocol request.** The budget information can be removed.

1. **Are any of the above funding source(s) a component of the Department of Defense?**

Yes  No

**If YES, please complete and append the** [**DOD attachments**](https://campusirb.duke.edu/forms) **to this protocol request.**

1. **Do you have an outside interest (financial or otherwise) that is in any way related to this study?**

Yes  No

**If YES, please explain.**

|  |
| --- |
| Click or tap here to enter text. |

|  |
| --- |
| Section 5: Research Question |

1. **What is your research question or the purpose of your research?**

|  |
| --- |
| The goal of this project is to quantitatively assess how prosecutors exercise their discretion during plea bargaining. In particular, this project aims to better understand both how the exercise of discretion varies *between* different prosecutors' offices, and also how the use of discretion in each office varies across defendants of different socioeconomic backgrounds. |

1. **Provide background information about the research that will help the reviewer understand your project.** Avoid discipline-specific jargon.

|  |
| --- |
| Despite making up less than 5% of the global population, more than 1 in 5 incarcerated people in the world are incarcerated in the United States. But this was not always the case — the roots of mass incarceration in the United States can be traced back to a four- fold increase in incarceration that began around 1975 and accelerated through the 1980s and early 1990s (Sawyer and Wagner, 2023).  One might naturally assume that this increase in incarceration was the result of a rise in violent crime, the War on Drugs, increases in mandatory minimum sentences, three-strike laws, or any one of the many other explanations that have captured broad attention over the past few decades. Increasingly, however, evidence compiled by scholars like Pfaff (2017) are pointing the causal finger in the direction of the often-underappreciated power of prosecutors. Data from the mid-1990s through the early 2000s, shows that the rise and incarceration cannot be explained by increases in crime, arrest rates, the share of felony cases that result in prison sentences, or in the length of sentences. Rather, it seems to be the case that the increase in incarceration appears to have been driven by a marked increase in the number of felony arrest charge filings by prosecutors.[[1]](#footnote-1)  The central importance of prosecutors in understanding mass incarceration in the United States is also reflected in plea bargaining statistics. Between 1980 and 2010, the share of adjudicated federal criminal cases that actually went to trial (as opposed to being resolved by plea bargain) decreased sevenfold, from 22.8% of cases to 3%. Similarly, a 2017 study of state courts found that more than 97% of criminal cases that were not dismissed were resolved by plea bargain in all twenty-two states studied (Hessick, 2021).  Yet for all their importance, how prosecutors exercise their substantial discretion what cases to pursue and what deals to make remain poorly understood. Unlike court proceedings — which are almost entirely public — the deliberations of prosecutors and the plea bargaining process itself are not publicly reported and are consequently difficult to study.  The goal of this project is to open the black box of prosecutorial decision-making by analyzing administrative data on the universe of North Carolina criminal proceedings initiated between the start of 2013 and the end of 2021. This public detailed data provides information on how cases evolve over time, including details of initial charges filed by prosecutors, whether cases are diverted or dismissed, the charges to which a defendant pleads guilty (if they plead), and information on the defendant and lawyers involved.  By studying how cases evolve from initial filings (which generally reflect the most serious crimes for which the defendant could potentially be convicted) to final resolution (dismissal, the charges to which the defendant pleads guilty, or trial outcomes), this project aims to systematically study variation in dismissal decisions and plea bargain leniency. Moreover, by pairing this with data on defendant characteristics (race, age, gender, whether they have a public defender or private attorney) and prosecutor characteristics (demographics as well as information on how their District Attorney has positioned themselves in DA elections), this analysis also aims to better understand how different factors impact the exercise of discretion. |

|  |
| --- |
| Section 6: Secondary Data |

Research involving the study of existing data may include individual records (e.g., academic, medical, financial), data sets, interview notes, biospecimens, online profiles and posts (e.g., social media), and audio- or video-recordings. These data could be available for purposes other than research, and may be provided with or without direct or indirect identifiers, including a key.

1. **Will any of your data be provided by the North Carolina Educational Research Data Center (NCERDC)?**

Yes  No

**If YES, append all the completed NCERDC forms to this protocol.**

1. **Do any of the data include (check all that apply):**

Medical records provided by Duke Health (clinic, department, or facility)

Medical records provided by a non-Duke entity

Academic records

Data provided by a component of the DOD (Department of Defense)

None of the above

1. **Identify the dataset(s).** Include the estimated number of records you will receive in each set.

|  |
| --- |
| The primary dataset is the Automated Criminal/Infractions System (ACIS) Criminal and Infraction Statistical Extract which contains criminal and infraction case information as recorded by the clerks of superior court in the 100 clerks’ offices in North Carolina for the 2013 and the end of 2021. All criminal defendant data will be anonymized by Sean Chen of the Duke University Law School Library prior to being provided to researchers. |

1. **Identify the individuals and/or organizations providing the data.**

|  |
| --- |
| Sean Chen (Data Services Librarian and Assistant Director for Metadata and Cataloging) at the Duke University Law School Library. Please see Appendix A for attestation that Duke Law Library has permission to provide the data to our team, Appendix B for the full email exchange leading to the initial procurement of the data by Duke Professor Michele Easter from the North Carolina Judicial Branch prior to deposit of the data with the Duke library system, and Appendix C for confirmation of deposit with the Duke Library. Although the contents of this data are a matter of public record, the Duke Law Library has developed a protocol for anonymizing criminal defendant and witness data to mitigate sensitivity concerns and improve ease of researcher access. This project will use the version of ACIS data with all criminal defendant and witness information anonymized. |

1. **Are the data publicly available?** Data are considered public if they are readily available for research purposes without making a formal request. That is, anybody can download the data with a simple click from an open, public-facing website.

Yes  No

1. **Describe the process or mechanism to obtain the data.** For example, you may need to create an account, register a study, submit an application, or enter into an agreement (e.g., a data use agreement or DUA). The process or mechanism could also include a personal arrangement with a colleague to provide you with the data.

|  |
| --- |
| Sean Chen (Data Services Librarian and head of Cataloging and Metadata Services) at the Duke University Law School Library has agreed to provide Nick Eubank with a version of the data with all criminal defendant information anonymized after receiving confirmation of IRB approval for our project. Appendix A of this application includes a letter from Sean Chen attesting to their willingness to provide this data, and Appendices B and C attest to how they got it. |

**Include in the Appendices any documentation that explain or describe the mechanism for obtaining the data.** Documentation may include an agreement, copies of confirmation emails from the data provider, or screenshots of a completed online data request application.

**Important items to note:**

* If the mechanism to obtain the data specifies requirements for how Duke must securely store or protect the data, the data meet Duke’s “sensitive” [data classification standard](https://security.duke.edu/policies-procedures-and-standards/data-security/data-classification-standard/). The “Sensitive” Data Classification questions (Section 7) must be completed.
* If the mechanism to obtain the data requires an institutional signature, researchers may not sign on behalf of the university.

1. **Describe the variables.** The description should include any direct identifiers (names, email addresses, images, or home addresses) and/or indirect identifiers (data points that, when combined, would allow someone to deduce the identity of the subjects).

|  |
| --- |
| The original (identified) ACIS data includes all administrative records related to criminal cases in North Carolina. This includes, for every case updated during the period in question, details of all charged offenses, criminal defendant name, aliases, birthdate, address, witnesses, DMV infractions, judgements, penalties, sentences, assigned judge(s) and attorneys. The data includes details of ~30,000,000 criminal cases.  All criminal defendant and witness data will be rigorously anonymized by Duke Librarians and all direct and indirect identifiers will be removed before I receive the data. However, it is worth reiterating that all information in this data — including identified criminal defendant data — is a matter of public record. The data is being anonymized here just to reduce any concerns about sensitivity as identified defendant data is not needed for the research being undertaken. The only identifying information that will be left in the data pertain to public officials and officers of the court (e.g., the presiding judge, prosecutors, defense attorneys, etc.) — information which is already a matter of public record. |

1. **If the data include any identifiable information (direct or indirect), will the identifiers be removed from the data either before or after you receive them?**

Yes  No  N/A - data do not include identifiers

**If NO, explain why the identifiers will not be removed.**

|  |
| --- |
| Identifying information on officers of the court is both already a matter of public record and is also required to undertake analyses like linking prosecutors or defense attorneys to electoral data on District Attorney races, to allow them to be identified if they move across jurisdictions, and to allow for statistical inference on attorney race and gender (using statistical models on differences in name use by race and gender in the United States). |

**If YES, describe the process for removing the identifiers, including when they will be removed and by whom.** In some cases, a third-party may be required to remove identifiers from the data.

|  |
| --- |
| All direct and indirect criminal defendant and witness identifiers will be rigorously anonymized/removed by Duke Librarians the data is provided to our team. The only identifying information that will be left in the data pertain to public officials and officers of the court (e.g., the presiding judge, prosecutors, defense attorneys, etc.) — information which is already a matter of public record. |

1. **Would an inadvertent release of identifiable data place individuals at risk of harm?**

Yes  No

**Please elaborate.** For example,explain why you are confident they are no risks, or if risks are possible, describe the risks and how they may be mitigated.

|  |
| --- |
| All records in this data are a matter of public record and are available in identified form from the North Carolina Judiciary. |

Note: If an inadvertent release of identifiable data may place individuals at risk of harm your data meet Duke’s “sensitive” [data classification standard](https://security.duke.edu/policies-procedures-and-standards/data-security/data-classification-standard/). The “Sensitive” Data Classification questions (Section 7) must be completed.

1. **Describe how Duke will receive the data.** For example, Duke researchers may need to download the data via a secure FTP service. If data will not be transferred to Duke please explain. For example, a data provider may require that Duke researchers remote into their internal servers to access the data.

|  |
| --- |
| Duke Law School library will share the data via a private SharePoint or Box folder. |

1. **Where will data be stored when they are “at rest” (i.e., not in use)?**

|  |
| --- |
| Duke Law School library has a designated storage for the dataset. When in use by our team, data will be stored on a private Github repository protected by a strong password and two-factor protection (a FIDO yubi-key) to facilitate collaboration, code reviews, and collaboration. |

Please review the [Developing Data Protection Plans](https://campusirb.duke.edu/node/96) guide for a list of best practices and recommendations from the IT Security Office (ITSO).

Note: Duke University has determined that [the use of non-Duke cloud services, like Google Drive, Apple iCloud, and DropBox are not approved](https://security.duke.edu/policies/duke-services-and-data-classification) for official Duke use.

1. **Where will the data be analyzed?**

|  |
| --- |
| The project team will analyze the data locally on their laptops, all of which have encrypted drives and are password protected. The laptops will have regular software updates enabled, anti-virus software (Duke supports Symantec) and a password-protected screensaver (per the data protection plan guidance). |

Please review the [Developing Data Protection Plans](https://campusirb.duke.edu/node/96) guide for a list of best practices and recommendations from the IT Security Office (ITSO).

Note: Duke University has determined that [the use of non-Duke cloud services, like Google Drive, Apple iCloud, and DropBox are not approved](https://security.duke.edu/policies/duke-services-and-data-classification) for official Duke use.

1. **Who will have access to the data?** If any non-Duke researchers will have access to the data, please identify them.

|  |
| --- |
| Only the project team will have access to the data. |

1. **Do any of your data providers require that the data be returned or destroyed after you have completed your analysis?**

Yes  No

**If YES, by what date will your data be returned or destroyed?**  
Click or tap here to enter text.

1. **Please indicate whether your research requires that you re-consent participants for the secondary use of their data.** If you must re-consent participants, please explain how participants will be re-consented and include the consent process in the Appendices.

|  |
| --- |
| No. |

|  |
| --- |
| Section 7: “Sensitive” Data Classification |

This section should be completed if your study involves the collection or storage of data that meet Duke’s “sensitive” [data classification standard](https://security.duke.edu/policies-procedures-and-standards/data-security/data-classification-standard/). Data are considered “sensitive” if an of an inadvertent disclosure of the data would pose risk of harm to participants or Duke is required to protect the data.

* If an accidental release of the data will place participants at risk of harm, the data are classified as **sensitive**.
* If a data mechanism specifies how Duke must be securely store or protect the data, the data are classified as **sensitive**.

**Research oversight offices apart from the Campus IRB may need to conduct an ancillary review before the IRB can issue a protocol approval. Campus IRB staff will notify you in the event your research must undergo an ancillary review.**

1. **Do your data meet Duke’s “sensitive” data classification?** Data are considered “sensitive” if an accidental release of the data will place participants at risk of harm or a data mechanism specifies how Duke must securely store or protect the data.

Yes  No

**If YES, the “Sensitive” Data Classification section must be completed.** Please answer the remaining questions in this section.

### **If NO, SKIP to Appendices**. Click the arrow to the left of “If NO” to temporarily collapse the ITSO questions.

1. **Please describe the devices (laptops, tablets, mobile phones, etc.) that will be used to collect, transfer, store, and/or analyze data.**

|  |
| --- |
| Click or tap here to enter text. |

1. **Please describe how the devices identified in 7.2 (above) will be protected.**

|  |
| --- |
| Click or tap here to enter text. |

Note: IT Security (ITSO) has identified [minimum security standards](https://security.duke.edu/policies-procedures-and-standards/device-security/minimum-security-standards-laptops-desktops/) that include encryption of the mobile device or laptop, application of security patches, installation and regular updates of antivirus, and a password-protected screensaver.

1. Who is your departmental or unit IT contact?

|  |
| --- |
| Click or tap here to enter text. |

1. Who is responsible for data security, including upgrades?

|  |
| --- |
| Click or tap here to enter text. |

1. Sensitive data must be stored securely. Select the ITSO-approved environment where you will store and/or analyze the data.

Duke’s Box

Duke's Microsoft OneDrive

Duke's Qualtrics

Duke's Zoom

Duke University Protected Network (“PN”)

Other (please specify: Click or tap here to enter text.)

1. If data will not be stored on an ITSO-approved server (see 7.6), where will they be stored? Be specific.

|  |
| --- |
| Click or tap here to enter text. |

Note: Duke University has determined that [the use of non-Duke cloud services, like Google Drive, Apple iCloud, and DropBox are not approved](https://security.duke.edu/policies/duke-services-and-data-classification) for official Duke use.

1. **Please identify the individuals who will have access to the data and describe their role in the project.** If non-Duke individuals will also have access to the data, please clarify whether data access will be on their local storage or if they will remote in to a Duke server.

|  |
| --- |
| Click or tap here to enter text. |

**If data include direct and/or indirect identifiers (see question 6.8), question 7.9 must be answered.**

1. **How will access to the identifiable information be controlled?** For example, identify any individual(s) responsible for authorizing access to identifiable information.

|  |
| --- |
| Click or tap here to enter text. |

|  |
| --- |
| Appendices: Study Documents and Consent Processes |

In this section, please include any study documents (e.g. recruitment materials, survey/interview questions, measures and instruments, DUAs, etc.) and consent processes that you will use in your study.

Appendix A



Appendix B











Appendix C





1. After an arrest, prosecutors have discretion over whether to file charges or have the arrestee released (i.e., they may *decline* to pursue a case). That is how the number of felony cases filed can increase even if arrests for felonies do not increase. This is distinct from *dismissing* a case, which occurs when a prosecutor decides to stop pursuing a case *after* charges have been filed, and *diversions*, in which a defendant agrees to participate in an alternate judicial process (e.g., drug court) before conviction. [↑](#footnote-ref-1)