Evaluating a Mental Health Co-Responder Program in Burlington, North Carolina

Methodological Appendix

October 2025

Jamie Vaske, Senior Research Associate, Criminal Justice Innovation Lab, UNC School of Government

Anna Marion, Graduate Research Assistant, Criminal Justice Innovation Lab, UNC School of Government

Hannah Turner, Senior Project Manager, Criminal Justice Innovation Lab, UNC School of Government

Contents

Introduction

This Methodological Appendix is an accompanying document to the report Evaluation of a Mental Health Co-Responder Program in Burlington, North Carolina. Here we provide additional detail on the methodology used to conduct certain analyses in the report. Each section below references the page number in the body of the main report where the relevant results are discussed.

Findings

Program Scope

Identifying Calls for Service……..Report p. 5

To narrow down calls for service that were potentially eligible for Law Enforcement Crisis Counselor (LECC) intervention, we first had to separate calls for service from administrative events recorded in the CAD data. As explained in the main report, not every entry in the CAD data is directly related to a call for service from the public. The CAD data also include administrative notes and other events that are self-initiated by officers or the crisis counselor.

The raw CAD data included 864,074 entries from January 1, 2013, through November 1, 2024. Several data filtering steps were used to screen out CAD entries not related to a call for service (CFS). Each CAD entry is assigned a code according to the nature of incident being logged. Five of these "nature codes" were excluded from the data set as they were almost exclusively used to reference administrative notes: ADMINISTRATIVE DUTY, MESSAGE, 911 HANG-UP, DUPLICATE, and FOLLOW-UP.

The FOLLOW-UP code was used frequently by the crisis counselors starting in 2021 to document follow-up case management activities, like calling or visiting established clients. Some of these follow up activities might occur soon after a call for service if the officer has identified a person in need and the LECC is able to connect with the person quickly. However, we determined that this code did not reflect real-time co-responses, and, as such, it was appropriate to drop those entries for the purpose of this analysis. Instead, we relied on the case management data for clearer information on follow-up contacts and post-incident referrals in later analyses.

Entries for administrative or follow-up actions also tended to have no notes or very short notes. Accordingly, notes shorter than 75 characters (~7-8 words) were excluded from the counts of CFS. Finally, if there were multiple callers for a single incident, the duplicate calls were removed.

The goal of these criteria was to eliminate readily identifiable non-call entries while minimizing the number of calls excluded. Table A below shows the number of entries excluded by category. After exclusions, we were left with 543,352 CAD events that more accurately reflected calls for service.

Table A. CAD Entries Excluded from Call for Service Analyses**

Year	LECC Entries Excluded Due to Follow-Up Nature	Additional LECC Entries Removed Due to Short Notes	CAD Entries Excluded Due to Follow-Up Nature	Additional CAD Entries Removed Due to Short Notes	Additional CAD Entries Removed as Duplicates
2013	0	0	826	2,762	0
2014	0	0	908	2,837	0
2015	0	0	991	2,781	0
2016	0	0	1,044	2,620	0
2017	0	0	1,204	2,598	0
2018	0	0	1,149	2,393	0
2019	0	0	1,529	2,210	0
2020	13	2	1,958	2,306	0
2021	274	49	2,178	2,027	1
2022	422	134	2,249	1,972	2
2023	563	287	2,602	2,078	0
2024	338	94	2,261	1,686	1

**Table Note: LECC entries are a subset of overall CAD entries, so the last three columns of this table summarize all of the follow-up/administrative entries that were excluded.

Identifying Calls for Service with a Behavioral Health Component……..Report p. 6

After identifying calls for service, a combination of keyword matching and nature code selection was used to identify the subset of behavioral health related calls that were likely to be LECC-eligible. The following criteria were applied to determine which calls were automatically eligible and ineligible.

1. The following nature codes were excluded: AD, ANIMAL, INFO, and TEST.
2. Any record with LECC, LECC2, or YCC as the main element in the 'Calltaker', 'firstdisp', or 'primeunit' fields was included.
3. Any record with the nature code MENTAL or MCOMMIT was included.

To all the remaining calls, we applied a Large Language Model that was trained to identify 130 keywords and keyword combinations in the CAD notes field, which is a free text field where dispatchers, officers, and the crisis counselor all enter notes about the call. We chose this list of keywords based on consultation with the department and extensive manual review of samples to decide which words reliably indicated that the call was behavioral health related. The list of keywords is below in Table B.

For records containing these keywords, the model assigned a match score of 1 and a count of the total number of keywords that appeared in the note. After additional manual review, we chose to include all calls with a match score of 1 as part of the filtered dataset. We found that, consistently, even calls with only one or two keywords present could be reasonably categorized as LECC-eligible. In total, 25,774 calls across all years were identified as LECC-eligible.

Table B. Keywords and Keyword Combinations Targeted in CAD Notes

10-73	73 history	PTSD+suic att	IVC+harm herself	papers+involuntary commitment
behavioral	73 issues	CORE+mental	IVC+kill herself	papers+73
breakdown	73+ARMC	CORE+RHA	IVC+mental hospital	
crisis	73+10-73	CORE+LECC	IVC+CORE	
erractic	73+mental	CORE+IVC	IVC+psychiatric	
mental	73+LECC	CORE+crisis	IVC+psych	
psych	73+homeless	ARMC+involunary commitment	IVC+dementia	
suic att	73+group home	ARMC+10-73	IVC+PTSD	
Alzheimer	73+RHA	ARMC+IVC	IVC+involuntary commitment	
bipolar	73+schizophrenia	ARMC+voluntary commitment	IVC+73	
dementia	73+mental health	ARMC+mental health	Papers+IVC	
hallucinating	73+crisis	ARMC+harm himself	Papers+ARMC	
hullicination	73+bipolar	ARMC+harm herself	papers+mental	
manic	73+chapter 90	ARMC+depression	papers+10-73	
mental health	anxiety+depression	ARMC+group home	papers+crisis	
mobile crisis	anxiety+ARMC	ARMC+mental health	papers+mental health	
paranoia	anxiety+mental	ARMC+bipolar	papers+RHA	
paranoid	anxiety+RHA	ARMC+schizophrenia	papers+LECC	
psychiatric	depression+LECC	ARMC+schizphrenic	papers+group home	
psychotic	depression+anxiety	IVC+ARMC	papers+schizophrenia	
schizophrenia	depression+mental health	IVC+mental	papers+schizophrenic	
schizophenic	dpression+mental	IVC+10-73	papers+biploar	
LECC	depression+73	IVC+crisis	papers+harm himself	
RHA	depression+ARMC	IVC+mental health	papers+kill himself	
voluntary commitment	depression+ARMC	IVC+RHA	papers+harm herself	
involuntary commitment	depression+suicide	IVC+LECC	papers+kill herself	
harm herself	depression+harm himself	IVC+group home	papers+mental hospital	
harm himself	depression+harm herself	IVC+schizophrenic	papers+CORE	
kill himself	PTSD+ARMC	IVC+schizophrenia	papers+psychiatric	
kill herself	PTSD+mental	IVC+bipolar	papers+psych	
suicide	PTSD+IVC	IVC+harm himself	papers+dementia	
panic attack	PTSD+suicide	IVC+kill himself	papers+PTSD	

Program Outcomes

Use of Force ……Report p.15

We examined Use of Force and Display of Force data provided by the Burlington Police Department from 2019 to 2024. The most direct indicators of mental health involvement in these data are separate yes/no questions for whether the citizen or officer(s) involved identified that the subject had a mental health issue. Additionally, the officers report various "citizen behaviors" that may be relevant to crisis intervention.

The analyses in the report distinguish a subset of incidents with documented citizen behaviors that potentially indicate a mental health concern, even if there was not a mental health issue identified by the subject or the officer. The classification of relevant citizen behaviors is detailed in Table C.

Table C. Officer-Identified Behaviors in Use of Force/Display of Force Incident Data Categorized as Possible Behavioral Health Indicators

Documented Behaviors Classified as Indicating Possible Behavioral Health Concerns	Documented Behaviors NOT Classified as Indicating Possible Behavioral Health Concerns
• Bizarre, unusual • Disorganized speech/communication • Disorientation, confusion • Hallucinations, delusions • Hopeless, depressed • Mania • Neglect of self-care • Out of touch with reality • Suicide behaviors, threat, attempt • Unusually frightened or scared	• Belligerent, uncooperative • Disorderly, disruptive • Other

Criminal Justice Outcomes

Court System Involvement……Report p.18

Matching Case Management and Court System Data

To determine which LECC clients had a history of court system involvement, we merged statewide court records data (from January 1, 2014 to December 31, 2023) to the LECC case management data (from January 1, 2021, to November 1, 2024).[1] We used probabilistic record linkage following the Fellegi-Sunter model as implemented in fastLink.[2] The merge was based on standardized full names, calculated age as of December 31, 2024, and gender.

Each court record was grouped by name, date of birth, and gender to assign unique IDs. Names were standardized to lowercase, punctuation was removed, and whitespace was trimmed. For the LECC case management data, row IDs were added to facilitate back-matching. We compared cleaned names using both Jaro-Winkler and cosine similarity measures, with high thresholds for strong and partial agreement (0.93 for

[1] This merge was executed by Ted Enamorado, Associate Professor of Political Science, Washington University in St. Louis.
[2] Enamorado, T., Fifield, B., & Imai, K. (2019). Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records. *American Political Science Review*.113(2):353-371. https://imai.fas.harvard.edu/research/files/linkage.pdf.

agreement, 0.85 for partial).[3] Age was compared by absolute difference within a very tight threshold (0.001 years). Gender was compared by exact match on the first character.

Following the Fellegi-Sunter model, we estimated match probabilities for each pair using an EM algorithm with specified priors for matches and non-matches. We considered any probability over 0.85 a match. For individuals who had multiple possible matches in the court records data with a probability at 0.85 or above, we randomly selected one of the possible matches.

We also conducted additional optimization when manual review discovered that fastLink alone did not identify all possible matches with a high probability. This involved:

1. Marking those with a match rate of 0.85 and higher per fastLink.
2. Identifying if the last name, first four characters of first name, and date of birth in the case management data matched the same combination (last + first 4 first name + dob) in the court records data.
3. In instances where step two 2 *did not* match, individuals were linked based on: (a) whether they were linked via fastLink, and (b) on whether they ever had a zip code in Alamance County or neighboring counties—Guilford, Randolph, Chatham, Orange, and Rockingham—per the court records data.

After the match, we identified 355 LECC clients with any history of court system involvement between January 1, 2014 and December 31, 2023.

Creating the Non-LECC Proxy Comparison Group

To provide a comparison point between LECC clients and other similar individuals, we created a matched proxy comparison group. LECC clients were matched to individuals within the larger court records data based on six variables:

- Gender;
- Age category;
- Race;
- Percent of charges initiated or served in Alamance County;
- Whether any prior convictions were for violent offenses per the Lab's categorization methods[4]; and
- Whether the individual was ever recorded as homeless.

The data were then divided into two separate datasets: a dataset with only those who used LECC services at least one time (dataset #1), and a dataset of individuals that did not have any LECC events recorded for them (dataset #2).

The initial matching of those using LECC services to those not using LECC services was conducted using the following steps:

[3] Enamorado, T., & Kaufman, A. (Forthcoming). The Power of Aggregation for Probabilistic Record Linkage. Washington University in St Louis.

[4] Additional information on how the Lab offense categorization methods can be found in the Methodology tab of our Measuring Justice Dashboard.

1. Matching the LECC users (dataset #1) to those who did not use LECC (dataset #2) based upon exact matching on gender, age categories, race, percent of charges in Alamance County categories, whether the defendant had a prior conviction for a violent charge, and whether the individual was ever recorded as homeless. Dataset #1 was then matched to the eligible comparison data (dataset #2) using a joinby command in Stata. There were two LECC users who did not match to an individual in the comparison group.
2. Given the many to many match, there were some LECC users who matched to only one comparison individual, while others matched to thousands of comparison individuals. To select a one-to-one match for those with multiple matches, we generated a random variable, sorted the random variable for each personid, and then randomly selected a match (match_tag) for each personid. That c_personid (comparison personid) that is flagged with a "1" becomes the match. A new variable (match_id) is created to identify the pairs. This created a 1:1 matching approach.
3. The file is then reshaped to be long and the court records data is merged in for each personid.

Before matching, 355 LECC clients had a criminal charge between January 1, 2014, and December 31, 2023, and there were 4,321,892 remaining individuals in the statewide court records data that could be potential matches. After exact matching on the six variables, there were 353 LECC individuals compared to 353 comparison subjects.

Table D below shows the distribution of the matching variables before and after matching for the LECC and comparison subjects. When compared to the full population of individuals in the court records data, the LECC group was:

- Less likely to be male (55.77%) than eligible comparison subjects (61.10%, z = -2.04, p = .04).
- More likely to be identified as Black (40.85%) than eligible comparison subjects (29.28%, z = 4.79, $p \leq .001$).
- More likely to be homeless (8.17%) than eligible comparison subjects (0.80%, z = 15.22, $p \leq .001$).
- More likely to have 75% or more of all charges served in Alamance County (50.42%) than eligible comparison subjects (1.42%, z = 13.82, $p \leq .001$)
- More likely to have a violent conviction between the years of 2014 and 2020 (14.93%) than eligible comparison subjects (2.61%, z = 16.14, $p \leq .001$).

The far-right columns of Table D below show that the comparison subjects are equal to the LECC clients after exact matching.

Table D. Comparison of Background Variables for LECC Clients versus Comparison Subjects Before and After Matching

	Before Matching		After Matching	
	LECC	**Comparison**	**LECC**	**Comparison**
Number of People	355	4,321,892	353	353
Percent Male	55.77%	61.10%	55.81%	55.81%
Percent Homeless	8.17%	0.80%	7.65%	7.65%
Race				
Percent White	50.70%	50.72%	50.71%	50.71%
Percent Black	40.85%	29.28%	40.79%	40.79%
Percent Hispanic	5.92%	14.25%	5.95%	5.95%

	Before Matching		After Matching	
	LECC	Comparison	LECC	Comparison
Age				
Percent Under 17	3.66%	2.00%	3.68%	3.68%
Percent 18 – 22	10.14%	9.43%	10.20%	10.20%
Percent 23 – 27	13.24%	14.09%	13.31%	13.31%
Percent 28 – 32	15.49%	14.68%	15.58%	15.58%
Percent 33 – 37	14.65%	12.66%	14.73%	14.73%
Percent 38 – 42	9.86%	10.98%	9.92%	9.92%
Percent 43 – 47	12.11%	9.15%	12.18%	12.18%
Percent 48 – 52	7.04%	8.09%	6.80%	6.80%
Percent 53 – 57	5.35%	6.48%	5.10%	5.10%
Percent 58 – 62	4.79%	5.15%	4.82%	4.82%
Percent 63 – 67	1.69%	3.38%	1.70%	1.70%
Percent 68 – 72	1.13%	1.98%	1.13%	1.13%
Percent 73+	0.85%	1.94%	0.85%	0.85%
Percent of Charges Served in Alamance County				
0%	20.00%	97.47%	20.11%	20.11%
1 – 25%	8.45%	0.46%	8.22%	8.22%
26 – 50%	10.99%	0.48%	10.76%	10.76%
51 – 75%	10.14%	0.18%	10.20%	10.20%
76%	50.42%	1.42%	50.71%	50.71%
Prevalence of Violent Convictions				
Percent with Violent Convictions	14.93%	2.61%	14.45%	14.45%

The analyses in the main report pertain to the prevalence and type of charges individuals have two years before and two years after their involvement with LECC. Identification of the charges before and after their earliest LECC referral date is straightforward for those that have LECC involvement. The comparison group does not have an earliest LECC referral date and thus their time period for calculating the number of charges before and after LECC involvement is not as straightforward. We chose to use the earliest LECC referral date as the anchoring date within a matched pair as the threshold for calculating criminal charges pre- and post-involvement.

Supplemental Analyses on Court System Involvement Using the Comparison Group

In addition to the analyses presented in the report, we conducted additional analyses on the highest charge for incidents incurred by LECC clients and by the non-LECC comparison group. 306 individuals (189 LECC clients, 117 comparison individuals) had a history of court system involvement in the two years before their initial LECC interaction.

As shown in Table E below, LECC clients with a history of court system involvement were more likely to have incidents where the highest charge was a violent misdemeanor or a non-violent misdemeanor compared to non-LECC individuals. LECC clients were less likely to have incidents where the highest charge was a traffic misdemeanor. For violent and non-violent felonies, the percentage point estimates for the LECC client group are higher, but they still fall within the confidence interval for the comparison group. Therefore, we cannot say there are meaningful differences between the two groups for those highest charge categories.

A chi-square analysis showed that the highest charge per incident significantly differed between LECC clients and non-LECC clients ($\chi^2 = 62.47$, $p \leq .001$).

Table E. Breakdown of Highest Charge per Incident Occurring Prior to Initial LECC Interaction for LECC and Non-LECC Clients, 2014 – 2023

Highest Charge Type	LECC Clients		Non-LECC Individuals	
	Number of Incidents	Percent (95% CI)	Number of Incidents	Percent (95% CI)
Violent Felony	18	4.33% (2.74-6.76)	4	1.93% (0.72-5.04)
Drug Felony	11	2.64% (1.46-4.71)	6	2.90% (1.30-6.30)
Nonviolent Felony	36	8.65% (6.30-11.77)	15	7.25% (4.41-11.68)
Impaired Driving	20	4.81% (3.11-7.33)	10	4.83% (2.61-8.75)
Violent Misdemeanor	78	18.75% (15.27-22.79)	18	8.70% (5.54-13.38)
Nonviolent Misdemeanor	128	30.77% (26.51-35.38)	28	13.53% (9.49-18.90)
Traffic Safety Misdemeanor	41	9.86% (7.33-13.12)	51	24.64% (19.23-30.97)
Traffic Administrative Misdemeanor	84	20.19% (16.60-24.33)	75	36.23% (29.95-43.01)
Total Number of Incidents	**416**		**207**	

We also looked at the highest charge for individuals who incurred new charges within two years *after* their initial LECC interaction. These results were similar to the analysis above. As shown in Table F below, among those that incur a new criminal charge within two years after their LECC interaction, LECC clients are more likely to have incidents where the highest charge is a violent felony, a violent misdemeanor, or a non-violent misdemeanor. They were less likely to have incidents where the highest charge was a traffic misdemeanor. Again, while the LECC client percentage point estimates are higher for non-violent felony and impaired driving offenses, those estimates fall within the confidence interval for the comparison group. Therefore, we cannot say there are meaningful differences between the two groups for those highest charge categories.

A chi-square analysis showed that the highest charge per incident significantly differed between LECC clients and non-LECC clients ($\chi^2 = 85.21$, $p \leq .001$).

Table F. Highest Charge Breakdown for LECC Clients and Matched Comparison Individuals Who Incurred New Charges within 2 Years of the Earliest LECC Referral Rate

Highest Charge Type	LECC Clients		Non-LECC Individuals	
	Number of Incidents	Percent (95% CI)	Number of Incidents	Percent (95% CI)
Violent Felony	32	8.86% (6.63-12.27)	5	3.76% (1.57-8.73)
Drug Felony	17	4.71% (2.94-7.45)	6	4.51% (2.03-9.69)
Nonviolent Felony	27	7.48% (5.17-10.69)	7	5.26% (2.52-10.64)
Impaired Driving	9	2.49% (1.30-4.72)	1	0.75% (0.10-5.16)
Violent Misdemeanor	59	16.34% (12.86-20.53)	11	8.27% (4.63-14.33)
Nonviolent Misdemeanor	141	39.06% (34.14-44.20)	22	16.54% (11.23-23.86)
Traffic Safety Misdemeanor	18	4.99% (3.16-7.78)	38	28.57% (21.52-36.83)
Traffic Administrative Misdemeanor	58	16.07% (12.62-20.23)	43	32.33% (24.92-40.74)
Total Number of Incidents	**361**		**133**	

To provide additional context for these findings, we examined the specific violent offenses that LECC clients and non-LECC individuals were charged with. Table G below shows the most prevalent charges for LECC clients when the highest charge in the incident was either a violent felony or violent misdemeanor.

The most common violent felony was malicious conduct by a prisoner/throw, which is typically charged when an individual throws bodily fluids on a detention officer while in prison or jail—a behavior that can be associated with severe mental illness. The most common violent misdemeanor charge is violation of a domestic violence protective order.

Table G. Most Prevalent Charges for LECC Clients with Highest Charge Violent Felony & Violent Misdemeanor Incidents

Most Prevalent Charges in Highest Charge Violent Felony Incidents	Most Prevalent Charges in Highest Charge Violent Misdemeanor Incidents
Malicious conduct by a prisoner/throw (n = 7)	DV protective order violation (n = 20)
Assault LEO/PO serious injury (n = 5)	Simple assault (n = 16)
Assault physical injury LE/prob/parole officer (n = 3)	Communicating threats (n = 15)
Assault with a deadly weapon intent to kill (n = 3)	Assault on a female (n = 9)
	Misdemeanor stalking (n = 8)

Table H shows the most common charges for the non-LECC proxy comparison group when the highest charge in the incident was either a violent felony or violent misdemeanor. The most common violent felony charge is common law robbery. The most common violent misdemeanor charge is assault on a female.

Table H. Most Prevalent Charges for Non-LECC Individuals with Highest Charge Violent Felony & Violent Misdemeanor Incidents

Most Prevalent Charges in Highest Charge Violent Felony Incidents	Most Prevalent Charges in Highest Charge Violent Misdemeanor Incidents
Common law robbery (n = 3)	Assault on a female (n = 5)
Voluntary manslaughter (n = 1)	Communicating threats (n = 4)
Assault by strangulation (n = 1)	DV protective order violation (M) (n = 2)
Robbery with a dangerous weapon (n = 1)	Cyberstalking (n = 1)
	Assault by pointing a gun (n = 1)
	Assault with a deadly weapon (n = 1)

Overall, there are a very small number of incidents that involve violent offenses. As in the report, we emphasize that there is no information in the court records data that indicates if an individual has a history of behavioral health concerns. Because of the nature of the program, we can assume that many of the LECC clients have a behavioral health concern. We cannot assume the same about the comparison group. This is a limitation of the comparison group and should be considered when reviewing these results.

Sensitivity Analyses

The findings above related to court system outcomes were subjected to sensitivity analyses where we replicated the approach above and employed different matching methods. In particular, we estimated five different types of models (with seven separate analyses) in Stata 19 (S.E.):

1. Re-matching the samples using the same exact matching methods above but with different matches for the comparison group to determine if the results changed when LECC clients were matched to different comparison subjects (note: two separate sets of sensitivity analyses were estimated for this method);
2. Re-matching the samples using an exact matching approach but using a many to many (m:m) method (rather than joinby) to link the LECC clients to eligible comparison group on the six matching variables (note: two separate sets of sensitivity analyses were estimated for this method). This set of analyses was estimated to determine whether results were sensitive to the exact matching approach;
3. Re-matching the samples using a many to many approach, but using the following covariates: age, gender, race, percent of charges in Alamance County, presence of a violent conviction, and total number of charges resulting in a conviction between 2014-2020 (0, 1, 2, 3+) to determine sensitivity of findings to covariates;
4. A propensity score matching approach using a 1:1 matching method using the same variables employed in the exact matching approach used in the final analysis; and
5. A propensity score matching approach using a 1:4 matching method. For the propensity score matching, we used nearest neighbor matching with a 0.25 caliper and no replacement for matching comparison subjects (i.e., thus a comparison subject was matched only once to a treatment/LECC subject).

While there were variations in the results, especially for the types of incidents where the number of cases per cell could be under 10 observations, similar conclusions were reached regardless of the matching approach.

Criminal Incidents Prior to LECC Involvement

- LECC clients were more likely to have criminal charges in the two years prior to their LECC involvement (53%) relative to the matched comparison subjects (33.14% – 47.04%).
- LECC clients' criminal history involved more violent misdemeanors (18%) and nonviolent misdemeanors (30%) than the comparison subjects (6.79% – 11.59% violent misdemeanors, 10.69% – 16.74% nonviolent misdemeanors).

Criminal Incidents within Two Years After LECC Involvement

- Across all methods, LECC clients were more likely to have incurred new incidents in the two years post-LECC interaction (43%) than matched comparison subjects (21.63% – 35.49%).
- The prevalence of new incidents involving violent charges was higher for LECC clients than matched comparison individuals.
    - Violent felony incidents: LECC (8.8%) vs. matched comparison (0.76% – 3.76%)
    - Violent misdemeanor incidents: LECC (16%) vs. matched comparisons (2.24% – 10.34%)
- Additionally, LECC clients were more likely to be charged with new nonviolent misdemeanors (39%) than their matched comparisons (11.72% – 20.61%).

The full results of the sensitivity analyses are available upon request.

Jail Incarceration ……Report p.21

Matching Case Management and Jail Data

The process to match the LECC case management and jail data is very similar to the process used to match case management and court system data.[5] We merged each record from the Alamance County jail dataset with the LECC case management data using probabilistic record linkage following the Fellegi-Sunter model as implemented in fastLink.[6] The merge is based on full names, gender, and age. Each jail record was grouped by first name, middle name, last name, race, sex, age, and date of birth to assign unique IDs. Name fields were cleaned to remove noise and standardized to lowercase with extra spaces trimmed. For the LECC files, variable names were harmonized, date of birth was converted to age as of December 31, 2024.

As suggested by Enamorado and Kaufman, we compared cleaned full names using a combination of Jaro-Winkler and cosine similarity, with agreement levels defined by similarity score thresholds (0.92 for agreement in names, and 0.80 for partial agreement).[7] Age was compared using absolute value of the differences, with tight thresholds for similarity (e.g., 1 week for agreement and 4 months for partial agreement). Gender was compared by exact agreement on the first letter.

Matches were assigned probabilities and filtered using EM-estimated thresholds. For edge cases (different names but same date of birth and same name but different date of birth), we applied stricter agreement criteria and flagged borderline matches for manual review.

After the match, we identified 201 LECC clients with any history of incarceration in Alamance County between January 1, 2014, and November 1, 2024.

[5] This merge was also executed by Ted Enamorado, Associate Professor of Political Science, Washington University in St. Louis.
[6] See note 2.
[7] See note 3.