Appendix

Table of Contents

Appendix A: The Risk Need Responsivity (RNR) Program Tool

Background: the Risk Need Responsivity Simulation Tool

The RNR Simulation Tool, developed by ACE! and partners with support from the Bureau of Justice Assistance (BJA: 2009-DG-BX-K026), operationalizes the RNR framework. The toolkit assists criminal justice supervision agencies and treatment providers to determine the forms of programming that will be most effective in reducing recidivism and improving outcomes for their population. The tool is also designed to guide resource allocation and help jurisdictions identify service provision gaps. The RNR Program Tool for Adults, one portal of the toolkit, is a 60-minute online program self-assessment completed by staff from service provider organizations and agencies.

The RNR Program Tool's underlying algorithms use the information provided to: 1) classify programs into one of six program groups based on the self-selected primary target behavior: Severe Substance Use Disorder, Decision-Making, Self-Improvement and Management, Interpersonal Skills, Life Skills, and Punishment only/other; and 2) assess programs' adherence to essential features of effective programs to reflect the overall quality of the program (defined below). Programs that adhere more closely to evidence-based practices relative to the defined primary target behavior will receive higher scores. All programs then receive feedback in the RNR Program Tool output to improve the quality of the program in each of the six domains.

The tool is intended to help criminal justice agencies better understand available resources and create a system that is responsive to justice-involved individuals' specific risk-need profiles. Providers can use this information to make changes to their programs and improve client outcomes. Criminal justice agencies' planning departments can use this as decision support for planning and implementation. For this project, ACE! tailored the underlying RNR framework to the specific nature of programming in New York City, which included modifying program classifications and domain areas.

Classifying Programs with the RNR Program Tool

For this project, four main program groups that are directly linked to a recidivism reduction effect were developed, 1) Severe Substance Use Disorder (SUD), 2) Decision Making, 3) Self-Improvement and Management and, 4) Social and Interpersonal Skill Development. A description of each programming model and the reported effect sizes are described below. While a person may present with multiple needs, their primary program placement is established in only one area based on their severity of need. Also, it should be noted that the estimated effect size that we use is for a moderate to moderate-to-high quality programs; higher quality programs can generate slightly higher effect sizes. Finally, some individuals do not present with a need for any clinically-based programming, as they are not estimated to have any specific criminogenic needs (rather, they might have an identified need for education or employment). This category of programming need is described as "No Formal Programming/Case Management" which can include case management to address service needs.

Group A – Severe Substance Use Disorder

Individuals with Severe Substance Use Disorder (SUD) require intensive (daily) programming with high levels of structure that occur over a longer period of time due to the nature of their drug use and the patterns of recovery. Many also present with a co-occurring mental health disorder. Some of the common types of treatment for severe SUD include residential treatment, therapeutic communities (TCs), and intensive outpatient treatment (IOP). Individuals who use opioids or alcohol often benefit from medication assisted treatment (MAT), such as buprenorphine, naltrexone, or methadone, but the MAT must be accompanied by behavioral interventions that help individuals address the underlying causes and correlates and identify strategies for relapse prevention.

Crimesolutions.gov rates 16 programs and 4 practices as "effective" for treating substance use disorder (2018). Some of the effective interventions and practices include MAT, contingency management, drug courts, Multidimensional Family Therapy, and Trauma Affect Regulation: Guide for Education and Therapy (TARGET). Many curricula that address severe SUD are cognitive-behavioral therapy (CBT)-based.

Programs in this category fall under American Society of Addiction Medicine (ASAM) levels of care 2.1 (Intensive Outpatient Services) to 4 (Medically Managed Intensive Inpatient Services). Residential programs such as TCs should provide anywhere from 15-30 hours of treatment per week and should last for 9-16 months. Intensive Outpatient Services should include a minimum of nine hours/week. For example, participants attend three times per week for three hours per day. Dosage in Group A programs may also vary by program phase. For example, phase 1 is the most intensive with participants attending anywhere from three to seven days a week. The next phase may entail going to treatment for two days per week.

The estimated effect size is moderate with a d=.30. That is, it is possible that the average person who completes the 12-18-month program would have a reduced probability of recidivism of around 20-25%. Individuals with severe SUD often present with multiple additional needs or destabilizers. They often benefit from additional programming around cognitive restructuring techniques. Group A programs should include clinical staff and use evidence-informed curricula.

Group B – Decision-Making

Group B programming emphasizes cognitive restructuring to change maladaptive thinking and behavior patterns. Individuals in this programming group often have a number of lifestyle and cognitive errors that affect impulsive decisions and risky behaviors and should receive programming multiple times per week and have moderately high structure.

Programming tends to be cognitive behavioral therapy (CBT)-based. Some curricula that address decision-making include Thinking for a Change, Moral Reconation Therapy, Decision Points, and Counterpoint. Individuals placed in this category tend to be moderate-high risk with multiple needs. In addition to the CBT program, individuals may need additional assistance in other areas such as housing, social support, or mental health maintenance. The estimated effect size is moderate with a d=.30. That is, it is possible that the average person who completes the 12-18-month program would have a reduced probability of recidivism of around 20-25%.

Group C – Self-Improvement and Management

Group C programming helps individuals develop social and problem-solving skills to address mental health, mild or moderate SUD, and self-control to learn to self-regulate behavior, manage emotions and manage conditions (substance use, mental illness, coping skills, etc.). Individuals receive programming weekly or several times per month, but the level of dosage will vary based on whether the individual has a mental health disorder and the severity of the mental health disorder. Restrictions in these programs should be low. Individuals in this category may present with co-occurring mental health and substance use disorders. Dosage of Group C programs depends on the severity of the mental health issues. Individuals who receive programming in this category should receive an evaluation from a mental health professional for professional clinical judgement on program dosage.

SUD programs in this category fall under American Society of Addiction Medicine (ASAM) levels of care 0.5 (Early Intervention) to 1 (Outpatient Services). Individuals in Group C often need medication management in conjunction with therapy. Some common curricula that address SUD include the Matrix Model, Helping Women Recover, Helping Men Recover, Kathleen Carroll's CBT, and Cognitive Behavioral Intervention for Substance Abuse. The estimated effect size is small with a d=.20. That is, it is possible that the average person who completes the three- to four-month program would have a reduced probability of recidivism of around 10%.

Group D – Interpersonal Conflict Skill Development

Group D programs are those that provide structured counseling and modeling of behavior to reduce interpersonal conflict and develop more positive interactions, as well as to develop social and communication skills especially with peers and loved ones. Programming is typically infrequent (e.g., monthly), and there are few restrictions on liberty.

Assessing Programs with the RNR Program Tool

After the RNR Program Tool categorizes programs into the appropriate program group, the underlying algorithms provide a score based on the program's adherence to evidence-based practices revolving around the RNR framework. The scoring algorithms also consider the level of programming. A description of the essential features of each domain is discussed below.

Risk includes two main items: the use of a validated risk assessment and target of a specific risk level. The Risk Principle asserts that risk for recidivism can be predicted using a validated assessment instrument and service intensity should be matched to this risk (Andrews & Bonta, 2010). The Risk Principle suggests that dosage should vary by risk level and individuals of different risk levels should not be treated together.

Need includes three items: behaviors associated with offending, tools to assess needs, and secondary needs. The Need Principle suggests that if recidivism reduction is the goal of an intervention then dynamic needs should be targeted (Andrews, Bonta, and Hoge, 1990). Andrews and Bonta (2010) describe eight dynamic needs: history of offending, personality patterns, cognitive distortions, destructive associates, family or marital circumstances, school or work conditions, leisure and recreational activities, and substance use disorder. Programs that focus on criminogenic needs receive higher scores than those that focus on needs that are not as closely associated with continued criminal justice involvement. Programs that attempt to target too many needs (or target behaviors) are likely to water down their program effectiveness.

Responsivity refers to matching the correct type of programming based on modality/content/approach for targeting behavior to an individual based on his or her risk and needs profile (Crites & Taxman, 2013). The Responsivity Principle suggests that programs (and recidivism reduction efforts) will be more effective if they are consistent with individuals' learning style, use cognitive-behavioral or social learning techniques, and address characteristics of an individual's personality that affect amenability to treatment (Andrews & Bonta, 2010; Andrews, Zinger, et al., 1990).

Implementation refers to the manner in which a program is carried out. Fidelity to the program model has been strongly linked to program effectiveness (Andrews & Dowden, 2005; Lipsey & Landenberger, 2005). Key measures include: how the treatment is delivered; staff credentials; staff training, quality assurance measures; program evaluation; technical assistance received; and inter-agency communication.

Dosage is the total amount of treatment an individual client receives. This may include a number of different indicators: total number of clinical hours; duration of the program (in weeks); frequency (number of days per week, amount (hours per week); whether the program includes phases; and aftercare. Dosage should be matched to risk and need (Bourgon & Armstrong, 2006; Mitchell et. al, 2012).

Structure includes controls placed on the individual as part of the program. These controls are used to help constrain individual behavior and/or movements. Constraining the movement of individuals can include: electronic monitoring, curfews, or day programs to support treatment goals (Drake, Aos, & Miller, 2009; Padgett, Bales, and Blomberg, 2006; Pattavina, Tusinski-Miofsky, & Byrne, 2009).

Program Tool Data Collection Methodology

Between June 2017 to March 2018, ACE! worked with program administrators in non-profit service provider organizations throughout the five boroughs to complete the RNR Program Tool through in-person or webinar sessions. During these sessions, ACE! facilitated completion of the online assessment and provided support and clarification where needed. Providers were instructed to gather pieces of information about their programs, such as number and demographics of individuals served, programming delivery models, and staffing. For a complete list of information collected in the Program Tool see Appendix D. ACE! researchers made multiple attempts to contact programs to collect unknown or missing information to complete the Program Tool, but some programs were not able to be reached. For documentation of missing Program Tool data, reference Appendix E. One-hundred ninety-four (194) programs, most of which were not funded by MOCJ, completed the RNR Program Tool.

After completion of the standard RNR Program Tool, ACE! researchers reviewed each program's responses individually to ensure correct program category placement. ACE! further honed program component requirements for program grouping than the RNR Program Tool requires (see Table 1 below). This was done to ensure consistency across programs within programming groups. For documentation of programs that were moved, see Appendix E. ACE! was also developed new Program Tool Scales to test the reliability of each Program Tool Domain, which is described further below.

	Table 1: Minimum requirements for program group placement						
Program Category	**Self-identifies target behavior**	**Use of relevant assessment instrument (s)**	**Use relevant approaches and/or curriculum**	**Dosage ≥ 100 hours**	**Provides group and/or individual therapy**	**Clinical staff with education and experience**	**Makes referrals based on assessments, no direct services**
A – Severe Substance Use Disorders	X	X	X	X	X	X	
B – Decision Making	X	X*	X	X*	X	X	
C – Self-Improvement and Management	X	X*	X		X	X	
D – Social and Interpersonal Skills	X	X	X**		X		
G – Case Management	X	X					X

***Exceptions on a case-by-case basis with other program components taken into consideration**
****Program tool indicates that most programs in these categories utilize a curriculum, but not required for category placement**
*****Group E programs provide classes, groups, workshops, or individual meetings**
Purple=Highest Level of Recidivism Reduction Programming; Blue=Critical for Stabilization (Indirect Relation to Recidivism Reduction)

Program Tool Scale Scoring Methodology

To create these new Program Tool Scales, Program Tool data from 2013-2018, which included roughly 1,800 Program Tool entries from over 200 jurisdictions across the US (including NYC) were used to complete the dimensionality, item grouping, and scale reliability analyses. The Program Tool entries used in the analysis were completed, first-time assessments and possessed minimal levels of missing data. The purpose of this methodology was to improve the scales in the Program Tool to enhance data collection and feedback provision. Principal Component Analysis (PCA) was used to identify items within the Program Tool that grouped together in a meaningful manner. Specifically, items were grouped together because they collectively measured the same overarching construct, such as case planning or use of rewards and sanctions. Based on the PCA results, several new constructs emerged within the Program Tool, which included different variables and more specific categories to enable ACE! to better focus the feedback provided to programs on how to improve their practices. Once the items were grouped together, a reliability test was conducted to verify that these items did indeed collectively measure the intended construct. Results from the reliability test indicated strong to excellent reliability, meaning that these scales do in fact measure their intended constructs. Table 2 below shows the Cronbach's Alpha value reliability for each program domain.

Table 2: Findings: Domain Cronbach's Alpha Value Reliability

<table>
<tr><th colspan="2">Domain</th><th>Domain Reliability</th></tr>
<tr><td colspan="2">Risk</td><td>0.807</td></tr>
<tr><td colspan="2">Need</td><td>0.762</td></tr>
<tr><td rowspan="6">Responsivity</td><td>Severe SUD</td><td>0.940</td></tr>
<tr><td>Decision making</td><td>0.861</td></tr>
<tr><td>Self-Improvement and Management</td><td>0.869</td></tr>
<tr><td>Interpersonal Skills</td><td>0.855</td></tr>
<tr><td>Life Skills</td><td>0.863</td></tr>
<tr><td>Case Management</td><td>0.874</td></tr>
<tr><td colspan="2">Dosage</td><td>0.814</td></tr>
<tr><td rowspan="5">Implementation</td><td>Case Management</td><td>0.971</td></tr>
<tr><td>Clinical Standards</td><td>0.838</td></tr>
<tr><td>Rewards and Sanctions</td><td>0.937</td></tr>
<tr><td>Quality Assurance</td><td>0.941</td></tr>
<tr><td>Drug Testing</td><td>0.919</td></tr>
</table>

The following changes within the Program Tool scoring domains incurred from the findings. Responsivity items were specific to the program group. The most weighted domain in the Program Tool—Implementation—was broken down into sub-scales representing the key areas of implementation measures. Finally, the domain—Structure/Restrictiveness—was removed due to a lack of reliability, and the drug testing items previously scored in this domain were included as a sub-scale measure of Implementation. Below is a brief description of the newly developed Implementation sub-scale domains.

Active Case Management Efforts*:* Case management practices that involve active case management to refer clients to services. "Active" refers to making phone calls, making appointments, working with service providers, etc.

Adoption of Clinical Standards: Program uses clinical standards including appropriate staffing, number of hours of programming, type of programming, and use of different clinical tools. Includes measures *of staffing patterns* to ensure the appropriate type of staff given the goals of the program, and *management of the program*, which entails having sufficient supervisors and leaders to manage the program.

Rewards and Sanctions: Program utilizes a system of structured rewards and sanctions. Rewards are based on targeted and relevant behaviors to the client's case plan/goals.

Quality Assurance and Fidelity: Has a process in place to measure that the program components and staff are following program procedures. Adherence to core concepts for a program in terms of key program components. Interagency Agreements are in place with other agencies to network services to ensure that clients have access to services.

Drug Testing: Programs that target substance use should utilize drug testing as a measure of success. Drug testing should be random.

Appendix B: Data Processing

Data for the research effort was obtained from several New York City agencies and service providers. These data were analyzed and combined to create a single analysis file that contributed to the offender level analysis. This chapter describes the data sources and the methodology used to combine information from multiple sources into the single analysis file.

We received data from three main sources—the Mayor's Office of Criminal Justice (MOCJ), Correctional Health Services (CHS) and individual service providers serving justice involved clients. MOCJ does not maintain its own data system but helped facilitate transfer of data from multiple NYC agencies and service providers. To mask the identity of individuals in the samples, MOCJ staff replaced all individual identifiers in all of the data files they provided us with a set of pseudo identifiers. The pseudo identifiers were constructed to preserve the ability to link records across multiple years and data sources.

Table A: Metadata for client-level data files received from MOCJ.

Data	Coverage	N	K
Summary Arrests (Detailed)	2009-2016	2,053,412	137
Summary Arrests (Limited)	2017	178,379	7
DOC admissions/discharges	2007-2017	919,445	301
ATI program participation	Up to March 2018	50,123	23
ICAN Assessments	2013-2015	11,215	56
CCI - I (Long)	2013-2014	906	235
CCI - II (Short)	2015	1,062	108

Data Received From MOCJ

The following data files were received directly from MOCJ. Table A provides metadata for the included files.

- *Summary Arrests Detailed* data (2009 – 2016): This file provided detailed information on all summary arrests in the five New York City boroughs between calendar years 2009 and 2016. Annual files are produced by New York City Criminal Justice Agency (CJA) and shared with MOCJ. They include client level detailed information (e.g., demographics), pretrial risk assessments (including several risk and needs related features), pretrial risk scores and recommendations, and criminal justice system related attributes (e.g., arraignment charges and, where available, disposition and disposition charges).The data includes all offenders formally arraigned in any of NYC boroughs but excludes offenders issued desk appearance tickets (DAT).
- *Summary Arrests Recidivism* data (2017): This data file included only summary arrests for 2017 (including charge and date) but did not include any of the additional detailed information available in the 2009-2016 file. This file was used only to compute recidivism measures through 2017 so that, for any member of the 2016 arraignment cohort, we could compute a 1-yr recidivism measure.
- *DOC Admissions and Discharges* (2007 – 2017): This was a single file including all DOC admissions and releases between 2009 and 2017. This data file included detailed information on offender attributes (e.g., demographics), and criminal justice status related attributes (charges, sentence lengths, and admission or discharge dates/types).
- *ATI Program Participants* (as of March 2018): This file included information on all offenders who were enrolled in an ATI program from its inception through March 2018. The data included enrollee

demographic information, ATI program name (and track), program start date and, if completed, program completion date and program exit reason.

- *ICAN* assessments (2013 - 2015): This file included risk/needs data on a sample of low-risk offenders screened for ATI program enrollment by the Independent Consumer Advocacy Network (ICAN).
- *CCI-I* long assessment (2013 - 2014): This file included a detailed questionnaire administered by CCI for the purpose of developing a risk/needs screener. The long assessment was given to a little over 900 misdemeanants, mostly post adjudication in Bronx Community Solutions, Midtown Community Court and Red Hook CJC. Interviews were conducted between May 2013 and February 2014. This was a purposive sample of high need, chronic misdemeanor offenders.
- *CCI-II* short assessment (2015): This file was a short version of the screener administered by CCI. The short assessment (CCAT-S) was given to a little over 1000 defendants held pre-arraignment in Brooklyn. Data were collected between May and December 2015

Each of the data file described above included one or more identifiers that were anonymized by MOCJ prior to transferring the data to the GMU/Maxarth team. These pseudo identifiers were used to merge records from all the data files into the main 2009-2016 Summary Arrest file. Each of the data files contained some elements of the risk and need features needed for our analysis as well as the ATI records for the recidivism analysis and quasi-evaluation.

While these data files provided enormous amounts of individual level details, because of the limited samples in several of the files (e.g., the CCI-I and CCI-II files) as well as the incomplete timelines in most (e.g., ICAN) three additional steps were taken.

1. The analysis was limited to just summary arrests in 2014, 2015, and 2016 (a three-year cohort). This was done, in part, because of the availability of data and, in part, so that analysis could be based a more recent cohorts.
2. Because the 1-to-1 matches using the pseudo identifiers provided relevant features for only a limited number of cases (e.g., using the CCI-I and CCI-II datasets), risk and needs flags were also obtained directly from service providers.
3. Despite the detailed client-level information available in the summary arrest records and the service provider records, specific chronic health problems—e.g., opioid abuse, serious mental illness (SMI), HIV/HCV, or other chronic illnesses—were not reliably available in these sources. A deidentified, random sample was requested and obtained from Corrections Health Services (CHS) and used to impute those additional flags into the main analysis file.

Data Received From Service Providers

There are numerous organizations providing services to justice involved individuals in New York City. The GMU/Maxarth team reached out to several of them and many volunteered to share client-level data with us for the analysis.[1] Our goal was to obtain risk and needs related data/assessments from these organizations so we may extract relevant information and use that information with our analysis. We obtained a total of a little under 100,000 assessments from a total of 11 service providers. Each of the participating service providers gave us data on risk and needs assessments that they may have conducted as well as client-level information (where available).

The data were a heterogenous mix of formats and structures. Our research staff scanned the data and extracted relevant demographic and risk/need features from each of the providers' data and combined them into a large assessment-level data file. While most of the providers had data at the client level, some provided multiple records for each client. To prevent overcounting these records, where data could be

[1] Because of contractual agreements between the participating service providers, GMU, Maxarth LLC, and MOCJ, we cannot identify the organizations who did or did not participate in the data sharing exercise.

collapsed to the client level, we did so aggregating risk/need indicators. For example, if the same client had a record in the mental health assessment data as well as in the substance use disorder assessment data, then we collapsed these two records into a single client level record with mental health and substance use disorder information included.

Figure A: Number of records in contributed data and availability of features in service provider data.

Data Sources	Records	Age	Ethnicity	Gender	Race	Prior Arrests (N)	Prior Convictions (N)	Prior Felony Convictions (N)	Prior Misd. Convictions (N)	Risk Classification	Criminal Thinking	Education	Employment	Family Related	Financial	Housing	Leisure Activities	Criminal Peers	Mental Health	Substance Use Disorders
All Data Sources	94,797	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Source A	924	X	X	X	X	X					X	X	X	X		X			X	X
Source B	1,098	X	X	X	X						X	X	X			X			X	X
Source C	18,730	X	X	X	X	X	X	X	X	X		X	X			X			X	X
Source D	11,215	X		X							X	X	X			X			X	X
Source E	9,296	X	X	X			X	X	X			X	X			X		X		X
Source F	140	X	X									X	X			X			X	X
Source G	692	X	X									X				X			X	X
Source H	5,044	X								X		X	X			X				X
Source I	257	X								X		X	X							
Source J	906	X										X	X							
Source K	46,495	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X

The service providers data, while very detailed, did not always provided information on the risk/need flags for every client. Figure A shows the availability of the various features in the service providers data. To extract the maximum amount of information from these data, a two-step procedure was used next.

1. All missing data was imputed within the service provider data so we could have complete client-level records to merge with the summary arrest data.
2. A profile-based strategy was used to merge completed records (original plus the imputed data) with the summary arrest file.

Each of these steps is described in more detail below.

Imputing Service Provider Needs Data

As Figure A shows, the service providers data had features from three common domains—demographic, risk, and needs. Most of the providers had good demographic data. A few had good, reliable criminal history and related risk measures. Most of them had good data on education and employment related needs, some had good data on substance use disorders and mental health, but very few had good data on criminal thinking, criminal peers, and financial stability.

In order to make the best use of the data, the first step was to fill in the missing information using all available data. The following steps were involved in this imputation process.

1. Estimate a series of multivariate models predicting each attribute on all other attributes. This included imputing missing demographic, risk, and needs features one at a time using all the remaining measures. For this first estimation round, all missing data on predictor features were coded as a separate category (for example missing gender was coded as a third category). That way all the data was used in this first round of imputing missing features (so no data was lost due to list-wise or case-wise deletion). At the end of this first round, the models were used to predict each of the missing values. With the exception of the criminal history measures, that were counts of number of prior misdemeanor or prior felony arrests, all of the remaining models were for categorical variables. Binomial or Multinomial logistic regressions were therefore estimated and a series of predicted probabilities were computed from these models.
2. Each of the predicted probabilities (for each model and each individual missing that feature) were next converted into predicted categories using the inverse transform method. The method is a universal strategy of sampling from any probability distribution using its cumulative distribution function and a random draw from the uniform distribution. Consider, for example, a client missing substance use disorder (SUD) need information. If the multinomial logit model using all other features predicted a 60% probability of the client not having any SUD need, 25% probability of having a substance abuse problem, and 15% probability of the client having a substance dependence problem, then the first step was to convert these probabilities into a cumulative probability function. The cumulative step function for this individual would have jumps at 60%, 85%, and 100%. Next, we generated a single random uniform number u between 0 and 1 and combined it with the cumulative probabilities to predict categories using the following logic: set SUD problem = “none” if $0 <= u <= 0.6$; set SUD problem = “abuse” if $0.6 < u <= 0.85$; and set SUD problem = “dependence” if $0.85 < u <= 1$. While each individual prediction is stochastic, the overall sample aggregates are guaranteed to follow the predicted probabilities. In other words, given that the randomly generated number u has the same chance of being anywhere between 0 and 1, the predicted category “no SUD problem” has a 60% chance of being selected; “abuse” has a 25% chance of being selected; and “dependence” has a 15% chance of being selected. This strategy was used to convert each and every predicted *probability* into a predicted *category*. These predicted categories were then passed on to the next iteration of models.
3. In the second modeling iteration, all the binomial and multinomial models were re-estimated but with the missing values imputed from round 1. Recall that in the first iteration, missing predictors were replaced with a separate missing category. In the second iteration, all missing values were replaced by their predicted categories from the first-round models. The second-round models were used to generate predicted probabilities and were then used to develop the second round of predicted categories using the inverse transform method again. These were then passed on to the next iteration.
4. This estimation/prediction/categorization/estimation loop was repeated a total of 5 times (including the first iteration). This resulted in 5 sets of predicted categories for each missing attribute for each person. The last three predicted categories were then combined to produce the final prediction—by popular vote. The most common prediction was set as the final imputation.

This procedure resulted in a client-level dataset with either raw or imputed values on all the variables shown in Figure A. These values were next fused with the summary arrest file.

Merging Service Provider And Summary Arrests Data

Consistent with the data sharing agreement with the service providers, their data did not contain any identifiers that could be used to link their data with the summary arrest files. As such, a different strategy needed to be developed to perform the merges. We used a profile-based strategy. Here, rather than use individual identifiers, all merging was done using *data hooks*—overlapping features in the summary arrest and service provider data files. The available data hooks included client age, race, ethnicity, gender, number of prior misdemeanors, number of prior felonies, risk-level, education level, and employment status. Merging was done over 14 different iterations. These are detailed below.

1. Merges in the first iteration were done using exact matches on all the data hooks items. I.e. a data point from the source data (service providers) was matched to a corresponding target (summary arrest) record if it matched exactly on all the data hooks listed above. If more than one source record were found with unique combinations of data hooks, the risk and needs flags were first aggregated. This resulted in a m:1 merge (i.e., each source record possibly merging with multiple target records exactly on all data hooks). Target records that were merged were set aside and the unlinked records were passed to the next iteration.
2. The same procedure was repeated by dropping risk level as a data hook but including all other data hooks. Unlinked records were passed to the next iteration.
3. The same procedure was repeated by dropping education level as a data hook and including all other data hooks. Unlinked records were passed to the next iteration.
4. The same procedure was repeated dropping employment level as a data hook and including all other data hooks. Unlinked records were passed to the next iteration.
5. The same procedure was repeated dropping employment and education levels as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
6. The same procedure was repeated dropping employment level, education level, and risk category as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
7. The same procedure was repeated dropping prior misdemeanors as a data hook and including all other data hooks. Unlinked records were passed to the next iteration.
8. The same procedure was repeated dropping prior felonies as a data hook and including all other data hooks. Unlinked records were passed to the next iteration.
9. The same procedure was repeated dropping prior misdemeanors and prior felonies as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
10. The same procedure was repeated dropping age as a data hook and including all other data hooks. Unlinked records were passed to the next iteration.
11. The same procedure was repeated dropping age and prior misdemeanors as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
12. The same procedure was repeated dropping age and prior felonies as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
13. The same procedure was repeated dropping age, prior misdemeanors, and prior felonies as data hooks and including all other data hooks. Unlinked records were passed to the next iteration.
14. In the final iteration, only gender and race were used as data hooks.

As the iterations enumerated above show, the merges get more and more fuzzy (or approximate) as the iteration number increases. Table B shows the proportion of linked records for each iteration. While the last few iterations are admittedly very broad, the proportion of links from these iterations are very small. In fact, more than half the links come from the most explicit links (iteration 1) and over 75% come from the first three iterations. The last 4 iterations only account for links to less than 1% of the summary arrest file. Indeed, the last iteration, using just race and gender, only results in links for 51 remaining cases not linked in prior iterations.

Table B: Profile-based record linkages between service providers and summary arrest data sources.

Iter. #	Data Hooks Included	Merged Target Records		
		N	%	Cumulat%
1	All hooks	369,510	55.25%	55.25%
2	All hooks, less risk	106,375	15.91%	71.16%
3	All hooks, less education	33,484	5.01%	76.17%
4	All hooks, less employment	22,556	3.37%	79.54%
5	All hooks, less education & employment	12,838	1.92%	81.46%
6	All hooks, less educ, empl, and risk	29,529	4.42%	85.88%
7	All hooks, less prior misd.	52,940	7.92%	93.79%
8	All hooks, less prior felonies	5,862	0.88%	94.67%
9	All hooks, less prior misd. and prior felonies	18,406	2.75%	97.42%
10	All hooks, less age	10,797	1.61%	99.03%
11	All hooks, less age and prior misd.	4,761	0.71%	99.75%
12	All hooks, less age and prior felonies	995	0.15%	99.90%
13	All hooks, less age, prior misd. and prior felonies	648	0.10%	99.99%
14	Only gender and race as hooks	51	0.01%	100.00%
		668,752		

Imputing CHS Data

The final source included in the analysis was data obtained from Correctional Health Services (CHS), New York City Department of Health and Mental Hygiene. Under a data sharing agreement negotiated between GMU, Maxarth, MOCJ, and CHS, a random sample of records (1 in 8) were shared with the GMU/Maxarth team. The records included client demographics, NYC borough of residence, criminal justice system related characteristics, and health related diagnostic information. CHS staff pre-processed the data to include flags for many health conditions including the chronic health conditions of interest for our analysis. A data file with K = 28 features for N = 7,000 client admissions from 2016 was provided to the research team. This represented a 12.5% random sample from all CHS admission for 2016 (~ 51K).

CHS data were merged with the summary arrest file (2014, 2015, and 2016 restricted version) using the same methodology as described above. Data hooks included all demographic, NYC borough, and criminal justice system related features. Because of the sparsity of the data, many-to-one links were generated using a single iteration. This meant some of the attributes needed to be collapsed into probabilities within data hook cells before the data merge. Once imputed into the summary arrest file, the inverse transform method (described above) was used to convert these probabilities into imputed categories.

Final Analysis File Contents

The above described strategy resulted in a final analysis file covering all summary arrests in NYC during 2014, 2015, and 2016. These records had detailed information from multiple sources either linked using available pseudo identifiers or imputed using profile-based links and stochastic imputations using the inverse transform method. The final data file included:

- Raw and Processed CJA data
  - Risk information
  - Charge information

    - Disposition information (where available)
- Linked DOC data
    - DOC admission dates
    - DOC admission types (city sentenced or detained)
    - DOC length of stay
- Linked Recidivism measures
    - DOC readmission
        - Dates (1-yr or 2-yr follow-up window)
        - Types (City Sentenced or Detained)
    - Summary re-arrests
        - Dates (1-yr or 2-yr follow-up window)
        - Types (Violent or Non-violent)
- ATI program data
    - ATI Program name/track
    - ATI Program start date
    - ATI Program completion date, if available
    - ATI Program completion status, if available
- Needs – obtained from multiple sources. The main needs used in our analysis were defined as follows:
    - *Mild/Moderate and Severe Substance Use Disorder*– This measure is created from numerous service providers data source. A 3-part variable is created that flags (i) No substance use, (ii) Mild/Moderate Substance Use Disorder, or (iii) Severe Substance Use Disorder. There are three ways these categories are defined:
        - Where Service Providers data includes a scale, the scale was converted into three categories.
        - Where Service Providers data includes 3 or more substance use groups, the groups were recoded into the three categories.
        - Where Service Providers data only permitted a Yes/No flag for substance use, models were developed to impute the type of SUD using all other need measures, demographics, and criminal history data.
    - *Opioid Use Disorder* – Obtained directly from CHS and imputed into CJA data. CHS definition is … "Opioid Use Disorder diagnosis: ICD9/10 codes (304.0#, 305.5#, F11.1#, or F11.2# where # can be other characters) + other confirmation (self-report, lab testing, clinical symptoms and assessment, medication orders)"
    - *Criminal Thinking* – Strong data not available from any sources. Instead Risk data used as proxy. CT need is flagged if the risk level = Very High. The underlying 5-category risk indicator was defined from a model of 1-year criminal recidivism using data gathered by CJA.
    - *Mental Health* – This measure is created from data provided by several service providers. There were four ways in which this flag is defined:
        - Where Service Providers data provided an assessment of MH, this was used directly.
        - Where Service Providers data provided detailed information for a screener (e.g., BJMHS – Brief Jail Mental Health Screener), information about the screener (found by online research) was used to create the MH assessment flag.
        - Where Service Providers data included MH assessment scores or totals from an instrument (e.g., PHQ – Patient Health Questionnaire), information about the screener (found by online research) was used to convert the scores into an assessment flag.

        - Where Service Providers data included ad-hoc questions that speak to the client's mental or emotional health issues, these questions were coded into MH assessment flags.
    - *SMI* – Obtained directly from CHS data and imputed into CJA data. CHS provided a flag for SMI.
    - *HIV/HCV* – Obtained directly from CHS data and imputed into CJA data. CHS provided a flag for HIV/HCV.
    - *Other Chronic Diseases* – Obtained directly from CHS data and imputed into CJA data. Defined by CHS as anyone with asthma, diabetes, hypertension, or seizure disorders.
    - *Housing* – This measure is created from several Service Providers data. This measure was created in one of three ways:
        - The Service Providers data indicated a housing need;
        - The Service Providers data indicated client was homeless;
        - The Service Providers data indicated that client provided homeless or transitional housing as current living arrangement.
    - *Education* – Created from self-report data collected by CJA. Defined as anyone without a high school, HS equivalent, or higher degree.
    - *Employment* – Created from self-report data collected by CJA. Defined as anyone who is not employed (either full-time or part-time).

Appendix C: Information Collected in the RNR Program Tool

- **Program Enrollment and Population**
  - Total number of clients served in last 12 months
  - Total number of criminal justice system clients served in last 12 months
  - Of those currently enrolled, age (16-27; 28-35; 36-42; 43+), race, and gender breakdown
- **Program Performance Metrics**
  - Percentage successfully completed in last year (overall and criminal justice population)
  - Percentage of how clients end program participation (e.g., successfully, dismissed, removal from program)
  - Of those who successfully completed, age (16-27; 28-35; 36-42; 43+), race, and gender breakdown
  - Percentage of program sessions completed by successful and unsuccessful clients
- **Program Staff**
  - Percentage of staff turnover in the past year
  - Percentage of employees in the organization who are contractors
  - Credentials of administrators and staff (e.g., advanced degrees, certification to teach a specific curriculum, relevant experience, correctional staff without certification)
  - Staff training and special qualifications (e.g., mental health qualifications)
- **Program Population and Eligibility**
  - Target population
  - Exclusionary criteria
  - Methods/ instruments used for screening clients for eligibility and matching clients to services, once enrolled
  - Criminal justice risk level for eligibility of the target population
- **Program Target**
  - Target behavior (e.g., substance use, cognitive restructuring, life skills)
  - Primary approach to target behavior (e.g., intensive outpatient, classes, case management)
  - Additional elements (e.g., family reunification, mentoring, housing, trauma-informed)
  - Flow and stages of programming and case/treatment planning (if applicable)
  - Curriculums used in the program
  - Referral practices to other programs
- **Program Controls, Sanctions, and Incentives**
  - Type of rewards/ positive reinforcement and sanctions/ negative reinforcement used, if applicable, and how they are administered
  - Frequency of drug testing in the program (if applicable)
  - Other types of structure/ controls
- **Program Dosage**
  - Total hours clients expected to complete in the program
  - Frequency of the programming (e.g., daily, weekly, monthly)
  - Hours per week participants are involved in the program
  - Length of the program, excluding aftercare length (in weeks)
  - Phases of programming
  - Components and length of aftercare
- **Program Implementation**
  - Requirements to complete program
  - Communication and information sharing practices with other agencies

- Prior evaluation of the program (e.g., external, internal, performance measures, client satisfaction scales)
- Program use of an operations manual or standard operating procedures (e.g., elements of manual used)
- Coaching practices with program staff (e.g., external coaching, peer/staff coaching)
- Quality assurance measures (e.g., external audits, videotaping sessions, staff review problem cases)
- Technical assistance received by the program in the last 12 months
- Contact methods used by program staff with clients
- Connections between clients and family (program with incarcerated populations only)
- Housing units (program with incarcerated populations only)
- Facilitation of reentry and programming by the facility (program with incarcerated populations only)

- **Program Funding and Expenditure**
  - Funding sources for the program, any fees-for-service

Appendix D: RNR Program Tool Documentation

Program Tool Missing Information

The following Program Tool question were not answered by one or more programs. Multiple attempts were made to receive missing information from programs.
*question is used in scoring.

- What is the zip code or zip codes in which the program operates? (Up to 5) (N=6)
- What principles of reentry do you adhere to? (N=3)
- In what setting does the program occur? Please select where the majority of the program takes place. (N=4)
- Please classify the target population for the program.* (N=1)
- Which of the following are exclusion criteria for your program? (N=34)
- How do you screen participants?* (N=3)
- What are the credentials of staff that run the program?* (N=9)
- Who administers or oversees the program? (N=1)
- Please estimate the percentage of employees in this program who are contractors: (N=13)
- Please estimate the percentage of staff turnover (defined as % of staff who have left the program in the past year): (N=13)
- What type of medical personnel work with the program? (N=6)
- Are the medical personnel linked to a Federally Qualified Health Center? (Don't know), (N=5)
- Are the medical personnel contractors or staff? (N=5)
- Please indicate the percentage of your staff with the following qualifications to address mental health issues. (N=18)
- What type of substance use does the program target?* (N=1)
- What substances are used by participants in this program?* (N=2)
- What medications are used in this program to treat substance use disorders?* (N=2)
- What approaches or methods are used to address this target behavior?* (N=4)
- What types of life skills does the program target?* (N=1)
- What types of interpersonal skills does the program target?* (N=1)
- How is drug testing administered?* (N=1)
- What is completion based on? (N=16)
- Of those currently enrolled in your program, please estimate the percent of individuals within the following age groups: (N=5)
- Of those currently enrolled in your program, please estimate the percent of individuals who identify with each racial/ethnic group. (N=12)
- Of those currently enrolled in your program, please estimate the percent of individuals who report being: (N=4)
- On average, what percent of all program sessions did the average client attend last year: (N=33)
- Estimated completion rate overall: (N=15)
- Estimated completion rate for criminal justice clients:% (N=17)
- Of those who successfully completed your program, please estimate the percent of individuals within the following age groups: (N=16)

- Of those who successfully completed your program, please estimate the percent of individuals who identify with each racial/ethnic group. (N=19)
- Of those who successfully completed your program, please estimate the percent of individuals who report being: (N=19)
- What is the intended duration of the entire program? (Do not include aftercare)* (N=12)
- How many hours per week is the participant required to attend programming?* (N=11)
- How often does a participant attend the program?* (N=5)
- Is aftercare required or considered a core component of the program?* (N=17)
- What type of services are part of aftercare: (N=1)
- What is the duration of required aftercare? (N=5)
- Is programming offered in phases?* (N=13)
- In addition to the primary intervention, what other elements does the program include? (N=5)
- Has the program ever been evaluated for its outcomes?* (N=1)
- How is/was the program evaluated?* (N=1)
- What techniques are used to ensure the program is of the highest quality (measure quality assurance)?* (N=2)
- Does this program use positive reinforcement and/or rewards?* (N=1)
- What type of positive reinforcement/rewards are used in the program?* (N=2)
- How do you notify participants about rewards that are given out?* (N=1)
- How are positive reinforcement/rewards given?* (N=8)
- Please select the ratio of positive reinforcers to negative reinforcers. (N=13)
- Does this program use negative reinforcement and/or sanctions?* (N=1)
- Please indicate on the line the ratio of controls to treatment used in the program. (N=4)
- How is information shared between correctional and program (i.e. treatment) staff?* (N=5)
- Does the program coach staff?* (N=2)
- What type of skills are staff coached on? (N=4)
- Has the program received technical assistance in the previous 12 months?* (N=3)
- What type of manual does the program use?* (N=6)
- What other types of contact do program providers have with clients? (N=3)
- How is this contact made? (N=21)
- From which of the following sources does the program receive funding. (N=6)
- How does your program bill for services? (N=2)
- When was the program first offered by the agency? (N=8)
- In what way has the program changed over time? (N=8)
- Please describe common issues you encounter when implementing the program. (N=12)
- How do you know the program is being delivered with fidelity? (N=14)
- What training or staff preparation is used for this program?* (N=12)