National Corrections Reporting Program (NCRP) White Paper Series

White Paper #3: A Description of Computing Code Used to Identify Correctional Terms and Histories

Revised: December 9, 2016

Prepared by:
Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Contents

1. Introduction

Abt Associates, in collaboration with the Bureau of Justice Statistics (BJS), has reoriented the National Corrections Reporting Program (NCRP) with the goal of improving its usefulness and reliability for describing and analyzing information about prison populations. We have done this by transforming the NCRP from a year-by-year accounting of prison admissions (A records), releases (B records), and prison stocks (D records) into inmate prison **terms**.[1] For individual offenders with multiple prison terms, we link them chronologically to assemble prison **term histories**. Thus, overall, for the period where states have reported to the NCRP, we construct a comprehensive dataset of prison term histories within states.

In addition, this reorientation has expanded data collection for offenders placed onto post-confinement community supervision (PCCS) after being released from prison. Previously, the NCRP collected only parole release records. Since 2011, collection has been expanded to include all admissions to and releases from any type of PCCS including, but not limited to, programs defined as parole supervision. Like prison records, these PCCS records have been transformed to assemble PCCS **terms** and PCCS **term histories**.

This NCRP white paper provides technical documentation for constructing prison and PCCS terms and term histories, as well as the construction of longitudinal histories linking prison and PCCS term histories together. With this documentation, NCRP data users will have the ability to understand, recreate and combine, if desired, prison and PCCS terms and term histories.[2] Thus, this paper provides details of the algorithm we use to process, characterize, and validate terms of incarceration and supervision reported to the NCRP. In the sections that follow, we discuss (1) how we identify and tentatively classify terms, (2) how we adjust and refine those tentative terms, (3) how we use D records to supplement the datafile, and (4) how we incorporate other adjustments into the final NCRP data.

We recommend that the reader review the second white paper (*Observations on the NCRP*) before continuing. That white paper introduces notation and concepts used in this current white paper. This current white paper revisits some of these concepts, albeit in greater detail.

---

[1] For extended discussion of changes to NCRP, see "National Corrections Reporting Program (NCRP) White Paper # 1: Observations on the NCRP."

[2] We also take additional measures to assess the data's general quality and reliability upfront. While we believe these steps to be crucially important in producing a reliable final product, they are not central in allowing researchers to reproduce the final datafiles, so we do not discuss them. As an example, we examine counts of D records over time to ensure that there are no anomalous changes in census totals. While this is an important diagnostic measure, no predetermined data manipulations are tied to the results of this output and its results alone do not per say affect the construction of the data.

2. Building Term Histories

2.1   A Global Process for Constructing Terms

Sections 2 through 5 that follow describe the various processes employed for constructing, validating and refining terms and term histories. There are three important features to this discussion. First, PCCS terms and term histories are built in exactly the same manner as prison terms and term histories; thus only one description of this "global" process is needed. Secondly, the data assembly and validation is complex and difficult to understand in the abstract. To make the discussion concrete, we describe the entire process in terms of prison records, however, the reader need only substitute descriptions of *prison* admissions/releases (i.e., A and B records) with *PCCS* admissions/release (i.e., E and F records) to fully understand the global process applied to both. Finally, it is important to note that while the prison data collection includes stock prison records (i.e., D records), there are no analogous stock records collected for PCCS records. As a result, sections that describe the use of D records to augment/refine prison terms (i.e., sections [3.2 – 3.4] & [4.1]) are not applicable to the construction of PCCS terms and term histories.

2.2   Establishing Individual Terms

The goal has been to reorient the NCRP from a year-by-year accounting of A, B and D records into terms and term histories. In theory, constructing term histories requires little more than linking chronologically ordered admission and release records. However, in practice this process is more complex. To explain the difficulty and how we overcome that difficulty, we start with a discussion of how we identify individual terms.

Some notation is essential. The notation A(1) → B(1,2) implies that the offender was admitted to prison on day 1 and was released from prison on day 2. The A represents an admission record, which has an admission date. The B represents a release records, which has an admission and release date. When there is no ambiguity, the notation A → B suffices. When the offender has multiple admissions and releases, these are represented with A(1) → B(1,2) … A(3) → B(3,4). The notation also covers more complicated patterns.

Our first step in identifying terms is to create a chronological ordering of an inmate's A and B records. To create this chronological ordering we use (1) inmate identifiers and (2) the date that each A and B record was recorded – admission dates for A records and release dates for B records.[3] From this ordering, we designate A and B records into individual terms by

---

[3]   For most states, we utilize the inmate identification number (inmate ID) where we have identified this as complete and reliable. In cases where state inmate identification number (state ID) is a more complete identifier, we opt to use this ID instead. For the 2015 data assembly, this included (CA, DE, MO, OR, PA, SC, TX)

grouping together records that share a common admission date.[4] The occurrence of a new admission date between records demarcates the beginning of the next term group. For example, consider an inmate who is admitted to prison (A), then released a few years later (B). After several years out of prison, this inmate is again admitted to prison (A), serves a subsequent term and then released (B). Chronologically, in this example, the inmate will have two term groupings by admission date: an A→B followed by a second A→B.

However, imprecise admission dates complicate grouping records. Consider a common illustration. Suppose we observe an inmate who is first admitted (A) and then released to parole (B), then later has his/her paroled revoked (A) and after serving the remainder of his/her sentence, is subsequently re-released (B). Logically, this sequence represents two distinct terms: [A→B followed by A→B]. However, if the admission date for the second B record were recorded as the *original* admission date, this sequence would be grouped as [A→B→B followed by A→.] using the admission date alone.[5] While using the original admission date to describe both release records may be a useful accounting method for correctional systems, it inhibits assembling a term history.

As the above problem illustrates, we cannot simply rely on admission dates alone to identify terms. However, we do not want to simply ignore this information either. Instead, we take additional steps to decipher what this information implies and ultimately integrate this information into our formulation of terms. By grouping together chronologically sorted records that share a common admission date, we create tentative terms that are known to be inaccurate but that are building blocks for identifying correct term records. From the example above, our solution would imply the identification of three separate terms: "A→B", followed by "A→.", followed by " .→B". This does not result in a set of terms that correctly express how we view this inmate's term history. However, it is *a first step* in the larger process of identifying and validating terms and term histories. From this point, we assess the validity of the proposed terms and make adjustments supported by the data. Before discussing the conversion of tentative term records into final term records, this white paper discusses some other preprocessing steps.

2.2.1 Correcting Inmate Identifiers

Occasionally the data indicate that a grouping of A and B records pertain to the same individual, but for some reason these records contain different inmate identifying numbers. We identify these cases by grouping together those records with the same reported admission dates, birthdate and sex for inmates, but *different* primary IDs. Then, within each group, we reassign records to have the same unique identifier. We use the identifier most commonly associated with the shared birthdate. In cases where more than one identifier is most

---

[4] The admission dates associated with the B records are the key ingredients to building individual terms. They tell us, for each release record, which admission record (and thus which term) that release should be associated with.

[5] The notation (A→.) denotes an A record that not followed by any record sharing the same admission date.

commonly associated with that birthdate, we choose the first identifier in the group (usually the smallest identifier value based on an ascending sort order). Overall, the need for correcting inmate identifiers is rare.

2.2.2   Records Created on the Same Day

Though rare, it is possible for multiple records to have been created on the same day for the same inmate. In cases where these are multiple A records, we retain the first record reported by the state and delete all others. In cases where these are multiple B records, we retain the B record with the earliest admission date, deleting all others. In cases where both A and B records exist, we use the admission date for the B record to assign a chronological structure; if the admission date for the B record occurs before the A record, we count the release as having occurred first. Otherwise if they share an admission date, we count the admission as having occurred first.

2.2.3   B Records with No Admission Date

In rare cases, B records are missing admission dates and cannot be linked to terms without imputation. In these cases, we use the admission date of the record that immediately precedes this B record in the chronological ordering of inmate records. We do this regardless of whether the preceding record is itself an admission or release record.[6] There are also some cases where no record precedes the missing information. In these instances we leave the admission date as unknown and thereby drop the term from the assembly.[7]

2.2.4   More Than Two Consecutive B Records

There are instances, though extremely infrequent, of terms with more than two consecutive release records that share a single, common admission date. The number of these consecutive records can range from three upwards. In a logical ordering of records, these terms appear as, A → B → B → … → B; however it is not totally clear why such a pattern might arise since it clearly defies the logical concept of a prison term. To create tentative terms, we collapse these multiple B records into a single term with one admission record and only two release records: the first and last observed release record. In other words, where there are more than two consecutive release records sharing an admission date, we delete the interior release records.

---

[6]   It is also possible that an admission record was created for the same day as the record with the missing admission date. In these cases, it is unclear whether the admission preceded or followed the release record. Our solution is to treat the admission record as subsequent to the record where we see that the inmate is incarcerated (according to the census (D) records) between that date and the date of the next recorded movement.

[7] This issue is rare in prison records but more common in PCCS records, especially in cases where PCCS records are generated from parole exit records previously collected for the NCRP. There are three states in which this occurs (NE, OR, and UT). In OR and UT, the date of release from prison is reported frequently enough to be used as a substitute. A fix is currently being tested for NE, but was not completed in time for inclusion in the 2015 annual archive.

2.3   Categorizing Term Sequences

Once we have established tentative terms (the "*first step*" from 2.1), our next step is to assess the quality of these tentative terms. Our goal is to assess the logical soundness of tentative terms and use suggestive evidence to improve the tentative terms. We do this by implementing a series of diagnostic measures coupled with decision rules to make changes where warranted by the data.

In order to determine if the established terms are sensible, we must look not only at the records that comprise each term, but also the records associated with terms that precede and/or follow a selected term. Consider, as a simple illustration, an individual term where an inmate has only a single admission record and no subsequent release record *sharing that admission date*. The table below depicts a simple conceptualization of this scenario.

Table 1

Term	Followed by…
A(1) [→ .]	???

The implication is that the offender entered prison on date 1 and has not left prison as of the most recently collected NCRP data. If this accounting is accurate, then the inmate should have no further records, A or B, and thus no further terms (i.e., the "Followed by…" box would be empty). If, however, we find that this inmate has additional terms following the admission record (e.g. a subsequent B record with a different admission date), then the accounting is imprecise and we must consider how this information affects our construction of term histories. Thus in order to examine the validity and reliability of individual terms, it is critical that we look at pairings of terms, or partial term histories, rather than terms in isolation from one another. We call these partial term histories "**term sequences**" and we classify every single term as fitting into one version of a term sequence.

Most term sequences conform easily to a consistent logical structure of admissions and releases. The simplest example is of those consistent terms that are not followed by any records whatsoever.

Table 2

Term	Followed by…
A(1) → . .	Nothing
. → B(1,2)	Nothing
A(1) → B(1,2)	Nothing

In addition to the three cases above, other cases also fit into a consistent logical structure. These cases are shown below.

Table 3

Term	Followed by…
. → B(1,2)	A(3) → . .
. → B(1,2)	A(3) → B(3,4)
. → B(1,2)	. → B(3,4)
A(1) → B(1,2)	A(3) → . .
A(1) → B(1,2)	A(3) → B(3,4)
A(1) → B(1,2)	. → B(3,4)

Where terms fit into term sequences that represent a consistent logical structure, we refer to these terms as **unambiguous.** That is to say that we consider the accounting of these terms to be fully accurate, such that no corrective action is needed. Some terms are considered unambiguous even where the larger term sequence is missing implied records. For example, the last row of the table above shows that an intervening A(3) record should be present in the sequence (it is implied by the B(3,4) record), though it is clearly missing. Nevertheless, because the A(3) record is implied by the B(3,4) record, and because this implied admission record is subsequent to the B(1,2) record, the term is considered unambiguous. We impute an A(3) record and classify this case as a valid, unambiguous term.

Other term sequences are illogical. We refer to these terms as **ambiguous**, meaning that constructing logical terms requires some corrective action. In short, term sequences with consecutive admissions (either as observed A records or as implied by B records) or sequences with multiple release records that share admission dates are ambiguous terms. The table below illustrates some common cases of ambiguous terms. For example, the third row of this table shows the case where an A record on date 1 is followed by another A record on date 2. Presumably there should be an intervening B record, however none is observed.

Table 4

Term	Followed by…
A(1) → B(1,2)	. → B(1,3)
. → B(1,2)	. → B(1,3)
A(1) → . .	A(2) → . .

A(1) → . .	A(2) → B(2,3)
A(1) → . .	. → B(2,3)

Resolving ambiguous terms in the overall construction of term histories is difficult in part because we cannot distinguish records that reflect true movements into and out of prison from those that do not. We have developed procedures for resolving these ambiguities by examining and comparing data consistency for these cases against that of unambiguous cases.[8] We discuss these procedures in the following sections.

2.4   Resolving Ambiguous Terms

In the white paper *Observations on the NCRP* we discuss the presence of ambiguous terms and the process for resolving them. The logical and empirical arguments for this approach are well documented in that source document, so we do not revisit the arguments for this approach in this section. However we do summarize the process for resolving ambiguous terms below.

Ambiguous patterns in term records take two general forms with many variations. One form is A(1) → B(1,2) → B(?,3) … The ellipses indicate that there may be more than two sequential B records. The question mark indicates that the second release record may or may not have the same admission date as the A record. Another form is … A(1) → A(2) → B(?,3) where the ellipses indicate that there may be multiple A records, and the question mark indicates that the admission date may refer to date 1 or date 2. A variation is … A → A → B → B … To resolve these term sequences, we do the following:

- Convert A(1) → B(1,2) → B(?,3) … to A(1) → B(1,K) where K is the last release date observed in the sequence, which may be after the observation window when B(1,K) is implied by a D record.

---

[8]   The method for resolving ambiguous sequences is provisional. A better method would use information about when offenders entered and exited community supervision. That information would allow us to insert an accurate A record into the sequence A→B→B when the explanation for the aberrant A→B→B sequence is explained by an intervening term of community supervision. Knowledge about entering and exiting terms of community supervision are not currently available to the NCRP.

- Convert A(1) → A(2) → B(?,3) to A(J) → B(J,3) where J is the admission date for the first admission record, which may be before the observation window, then A(J) is implied by a D record.[9]

There are two important notes about this approach. The first is that the entire problem is likely to be temporary. Beginning in 2011, we began collecting data on community supervision terms. The presence of community supervision data should help determine whether we are making the appropriate decision. The existence of an unambiguous community supervision term within the ambiguous sequence A → B → B would indicate that the sequence should be revised to A → B A →B, with a term of community supervision between the first B and the second A record. If cases where such a term does not exist, then we would know for sure that the sequence should be revised to A → B.[10] In short, we have adopted what appear to be the most reasonable rules for resolving the ambiguous sequences, and we expect to improve the rules for resolving ambiguities in the future. In a small number of exceptional cases we do modify this approach to handle cases (in 8 select states) where we have direct knowledge that ambiguous sequences are the result of recidivistic events. The details of this modified approach are found in the Appendix.

The second is that, the potential error introduced by this process is likely to be modest since most terms are unambiguous.[11] We note here that the Reference Guide in Section 9 provides a useful, high-level summary of the various terms we identify in the data, a range of the typical frequency with which these terms appear across states and an annotated description of the rule(s) for imputations we apply in these cases.[12]

---

[9] There is one exception. When we observe the sequence (A(1) → B(2, 3)), we compare the admission date on the release record to all previously observed admission records. If we observe a match to an earlier admission record for over half of the cases where the admission date falls in the observation window, we apply a different rule. In these cases, we ignore the admission date on the release record, thus combining the reported admission date with the release that follows it. The reason is that we believe the term is for a revocation to prison and that the original admission date on the release record refers to the initial admission to prison date.

[10] The use of community supervision terms is only beneficial if terms themselves are unambiguous. When community supervision data contradicts prison data, and both sources are data have the same degree of ambiguity (both ambiguous or unambiguous), we trust the prison data and move community supervision dates accordingly.

[11] As of now, evidence suggests that the proportion of unambiguous terms in a typical state is between 88 – 98% of all terms identified from the A and B records.

[12] The distributions reported in this appendix are subject to change as more data become available.

3. Refining Term Histories

Once we have identified, classified, and resolved terms using the decision making principles outlined above, we are left with fully described term histories. However, these histories contain inaccuracies, some of which were introduced by the earlier decision rules. We take additional steps to inspect term histories and make necessary refinements.

3.1 Overlapping Terms

Terms sometimes overlap. For example, consider two terms recorded as A(1)→B(1,3) and A(2)→B(2,4). This may be a legitimate sequence representing a *sentence* that starts on day 1 and ends on day 3, and a *separate sentence* that begins on day 2 and ends on day 4, however, for us this results in an inconsistent term history. Thus where terms overlap, corrective action is needed to construct consistent term histories.

Where we see multiple unambiguous terms overlapping, we collapse these terms together into a single (combined) term (see Section 5.0). However, if one of these terms is ambiguous, specific rules must be introduced to ensure that sensible steps are taken in light of possible imputations produced by earlier processes. There are two cases where we introduce refinements of the earlier processes:

- The case A(1) → B(1,2)…A(3)…B(1,4) becomes,

    A(1) → B(1,2)…A(3) → B(3,4), where an imputation is done for A(3)…B(1,4)

- The case A(1)…A(2)…B(1,3) becomes A(1) → B(1,3) where A(2) is dropped

3.2 Admission Observed Last

For many term histories, we observe the last record in a term history as an admission record with no observed release. Such a term implies that an inmate is currently incarcerated, and in the vast majority of cases this is true. Occasionally however, inmates who are observed to be currently incarcerated have actually been released and for some unknown reason they lack a release record indicating that release. Although rare, these anomalies introduce disproportionate biases into the term file. This is because inmates who are missing this final release record continue to be mistakenly identified as present in the prison population indefinitely into the future. The consequence is that over time, the size of the true prison population is progressively overstated.

To correct this false accumulation of inmates, we take the additional step of tracking D records for inmates who have not yet been released according to B record accounting, and examine the last reported set of D records to confirm the continued incarceration of those inmates. If an inmate appears in this set of records, we consider the term accurate and no imputation is needed. An inmate missing in the last set of reported D records implies the

release of that inmate since their last admission. Thus, we impute a release date. To do so, we search backward in time through the previous D records, up to the date of the last admission, in order to find the most recent D record reported for that inmate. Once identified, we impute a release date that randomly lies in between the date of the most recent D record and the date of the next set of D records. Explicitly, we impute a release date (d_B) for each inmate such that,

$$[3] \quad d_B = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$$

where,

r = a randomly generated number on the interval from [0,1],

d_{D1} = is the date of the latest observed D record for an inmate, and

d_{D2} = is the date of the next reported set of D records, following d_{D1}.

In the event that we find **no** D records after the last admission date, we simply impute a release date that is 30 days after the admission date.

3.3   Release Observed First

Often a release record with no observed admission record is the first observed record either because the admission took place before the reporting period or because of missing data. In most cases we simply take the admission date associated with the B record as the true admission date. However, in cases where the admission date is missing or inaccurately reported, we again must take corrective action to minimize distortions of prison population counts. On one hand, if the reported admission date comes *before* the true admission, then the presence of an inmate is reported falsely, leading to an <u>overcount</u> of inmates. On the other hand, if the reported admission date comes *after* the true admission, then the absence of an inmate is reported falsely, leading to an <u>undercount</u> of inmates. In this section we discuss our imputation for reducing the likelihood of overcounting inmates. For the discussion of how we deal with undercounting of cases, see the Section 4.1.3, *Peripheral D Records*.

To minimize distortions of population counts, we again take the additional step using D records to assess whether the admission date is supported by the data. We begin by looking at the first reported set of D records after the reported admission date.[13] If an inmate appears in this set of records, we consider the term accurate and needing no imputation.[14] If however an inmate is missing in this first set of D records, we search forward in time through the subsequent D records, up to the date of the first release, to find the earliest D record reported

---

[13]   For missing admission dates, we simply look at the first set of reported D records.

[14]   If an inmate is present in the first set of reported D records **and** they have a missing admission date, we cannot impute a reasonable admission date. Thus the start of their term remains treated as "unknown".

for that inmate. Once we have identified this record, we impute an admission date that lies randomly in between the date of the earliest D record and the date of the next set of D records reported Explicitly, we impute an admission date (d_A) for each inmate such that:

$$[4] \quad d_A = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$$

where,

    r = a randomly generated number on the interval from [0,1],

    d_{D1} = is the date of the earliest observed D record for an inmate, and

    d_{D2} = is the date of the next reported set of D records, following d_{D1}.

In the event that we find **no** D records before the first release date and after the implied admission date, we simply impute an admission date 30 days prior to the release date.

3.4   Partially Unreported Terms

Ambiguous term sequences arise for one of two reasons. The first is that the data reported by the state are complete, but logically inconsistent. For these cases we have discussed the decision rules implemented to resolve ambiguous terms. The second reason is that states sometimes have gaps in reporting records for some, but not all, years. The result is that omitted records create the appearance of ambiguous terms in the identification of term sequences. We discuss these cases and our approach for resolving them in this section.

Problematic terms manifest in two ways. The first is that some terms are only partially reported. There are two variations of this problem: (1) an A record appears with no B record; the B record was generated in a year where no data was reported, and (2) a B record appears with no A record; the A record was generated in a year where no data was reported. These partially reported terms are the subject of this section. The second manifestation is where no A or B records for a term are ever observed, because both the A and the B records were generated in years where no data was reported. We discuss these terms in the Section 4.0, and do not discuss them further here.

The best way to handle missing data is obvious: request the missing data from the states. The NCRP team has implemented this strategy. But when a state is unable to provide historical data, the solution requires imputation.

Dealing with ambiguous terms that result from partially reported terms requires that we introduce additional rules for imputation. Consider first the case of a B record with no accompanying A record. This case is the simplest and requires no new imputations, since the admission date is included as part of the B record. We simply use this admission date as the start date for this term and move on to the next term sequence. The case where we observe an A record without an accompanying B record is more problematic. We do not necessarily

want to treat this case as ambiguous, if we can determine that the release occurred in a year where B records were missing.

Again we use the D records to inform imputations. We identify those ambiguous cases where the release occurred in a year where B records were missing by looking at the D records before and during each gap in reporting, moving forward through time. Where we find that an inmate exits the prison population in one of these unreported years, we impute a release for them in the first year they exited. We impute this date using an expression similar to [3] from earlier, such that,

$$[5] \quad d_B = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$$

where,

 r = a randomly generated number on the interval from [0,1],

 d_{D1} = is the date of the latest observed D record before they exited, and

 d_{D2} = is the date of the reported set of D records in the year they exited, following d_{D1}.

3.5 An Alternative Imputation Strategy

The imputation strategies described in the previous three sections applies a uniform random number to assign an admission or release date when one is unknown. This should produce acceptable statistics because most terms are unambiguous and when the random number is applied, it should yield time-served estimates that are correct on average. A better procedure would apply imputations based on an analysis of time served. Then the admission date or release date would be imputed based on the time served analysis. We have not implemented this solution.

4. Supplementing Term Histories

The identifications, imputations and refinements discussed above have all involved terms that we observe either completely or partially through reports of A and B records as supplemented with D records. However there are other terms that we do not directly observe in any A or B records. These terms result for one of two reasons: (1) both admissions and releases occur during periods of time where the state does not report A and B records, i.e. the "gaps-in-reporting" problem, or (2) A and B records occur during periods where the state *did* report records, however these records are simply not reported. We call these **invisible terms**. In both of these cases however, the presence (or lack) of D records allows us to at least partially identify these terms. As a result we use D records along with reasonable imputations to augment term histories to include these invisible terms.

4.1   Invisible Terms with D Records

For some inmates, no A or B records are ever observed in the data, yet the D records show us that at least one term must exist. For other inmates, the D records reveal the presence of invisible terms that are part of a larger term history. Further, these invisible terms may lie at the beginning or end of a term history (we call these **peripheral**), or they may lie between observed terms (we call these **nested**). Each of these cases changes the way we think about the information in a term history, and thus how we handle the introduction of invisible terms.

4.1.1   Only D Records

For some inmates, we never observe an A or B record; term histories exist but are invisible. The most obvious case is an inmate who entered prison prior to the reporting period and remained in prison at the end of the reporting period. Naturally however, these inmates must have entered prison at some point, otherwise we would not observe them in the data (through the D records). We identify inmates with D records and no other recorded movements as those with invisible terms, and create a term history for these inmates using information from the D records.

We start by treating all of the D records for each inmate as part of the same initial admission (according to the first non-missing admission date observed in the D records), and thus part of the same single (invisible) term.[15] To impute admission dates, we first locate the earliest D record for an inmate to identify where they "enter" the data according to the D records. From this date, we apply the following rules:

1. If the earliest D record that we *actually* observe is the same as the earliest D record we *expect* to observe according to the admission date in the D record, then use the admission date as the term start date.
2. Otherwise, we ignore the reported admission date and impute the start of the term by choosing a date at random between the date of the earliest observed D record and the date of the preceding set of reported D records. Formally, we compute this imputation as,

---

[15] We argue that this is a sensible approach for several reasons. The first is that the admission dates recorded with the D records are of questionable reliability. It is not uncommon to for some states to report admission dates in D records that reflect the first admission in an inmate's entire history. Secondly, if multiple D records do show different admission dates, and this change reflects a true movement, then we should observe both a B record for the exit and an A record for the return. However, we argue that, in general, it is highly unlikely that a state would fail to report both of these events where a movement takes place. Finally, it's when the time span of reported records is very long that this assumption becomes questionable. However, we argue that the number of cases that fall into this group when the time span of records is lengthy is a trivial number. Inaccuracy in our assumption cannot significantly adversely affect the final term-level datafile.

$$[6] \quad d_A = (d_{D0} + 1) + (r)(d_{D1} - d_{D0})$$

where,

r = a randomly generated number on the interval from [0,1],

d_{D1} = is the date of the earliest observed D record for an inmate, and

d_{D0} = is the date of the reported set of D records immediately <u>preceding</u> d_{D1}.

To impute release dates, we locate the latest D record for an inmate and identify where they "exit" the data according to the D records. From this date, we apply the following rules:

1. If the latest D record that we *actually* observe is the last D record we can *possibly* observe, then we leave the term end date as ongoing (i.e., blank).
2. Otherwise, we impute a release date between the date of the last observed D record and the date of the next reported D record. This imputation is identical to the one used in equation [3] from earlier,

$$[7] \quad d_B = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$$

where,

r = a randomly generated number on the interval from [0,1],

d_{D1} = is the date of the latest observed D record for an inmate, and

d_{D2} = is the date of the next reported set of D records, following d_{D1}.

4.1.2   Nested D Records

Some inmates, like those above, have D records that stand alone, with no other associated term history. Other inmates though have D records that appear alongside other terms, but for which the logical structure of admission and releases in their term history does not appear to explain the presence of the D records. As an illustration, consider an inmate with an observed term history like the one below:

Table 5

	2002		2003		2004		2005	
	A(1)	D(·)	B(1,2)	D(·)	A(3)	D(·)	B(3,4)	

The D records located within A(1) → B(1,2) and A(3) → B(3,4) are accurate and sensible. However, another D record (in the middle) inexplicably appears between the B(1,2) and A(3) records. It is not clear what this extraneous record represents though it is clearly inconsistent with the logic of the reported records. We call D records that appear between terms, nested D

records.[16] We take steps to resolve these cases; however, we exclude cases where nested D records result from another imputation process used earlier. We do this to preserve the integrity and validity of the imputations already adopted.

In order to reconcile nested D records we must take one of two actions. One action is to impute an entire additional term (admission and release) that encompasses the nested D record(s). Another action would be to alter the release and/or admission dates of the existing terms so that they encompass the nested D record(s). Which action we take depends on whether the nested D records are reported consistently (or continuously) at every possible interval between the two terms. That is to say that, if we observe a nested D record each time it is possible to observe a D record between two given terms, then we believe that no time away from prison actually took place.[17] Our solution in this case is to combine the two existing terms by dropping the release from the first term and dropping the admission from the second term. So for the example above, since a nested D record appears every time it is possible for a D record to appear (only once in this example), we combine the individual terms A(1) → B(1,2) and A(3) → B(3,4) into one continuous term in prison, A(1) → B(1,4).

However, sometime we do not observe a nested D record for each set of D records reported on the interval between terms. Table 7 below shows an example of this:

Table 7

	2002		2003		2004		2005		2006	
	B(1,2)			D(·)		D(·)			A(3)	

In this table, D records do not appear on the intervals between 2002/03 and 2005/06. As a result we do not consider the prison stay to be continuous. In this case, we take the step of imputing an additional, intermediary term. The imputation process here follows the similar steps as for inmates with only D records, with one primary difference. In this case, D records are only evaluated on the interval between terms, rather than on the entire interval of all reported years. To impute term admission dates, we examine the admission date of the earliest reported D record on the interval between terms. Then,

1. if the admission date is equal to or earlier than the date of the B record for the first term, we drop that release record.

---

[16] We ignore any difficulties created by D records that occur on the same day as a release. We argue that this can reflect a consistent accounting.

[17] We argue that, in general, it is unlikely that a state fails to report both an admission *and* release given that both truly occurred.

---

2. otherwise we introduce a new term that start on the admission date reported in the D record. In the event that no admission date is recorded on the D record, we impute an admission date according to equation [6].

To impute release dates, we locate the latest D record on the interval. From this date, we impute a release date by selecting a date at random that lies between the date we are using and the date of either (a) the following D record or (b) the A record of the second term, whichever comes first. This imputation is identical to the one used in equation [7] from earlier.

Finally, if after the imputation process for creating new terms,

1. no possible D records can be observed between the B record for the first term and the admission date for the new term, then these terms are combined by dropping the B record and imputed A record altogether.
2. no possible D records can be observed between the A record for the second term and the release date for the new term, then these terms are combined by dropping the A record and imputed B record altogether.

4.1.3   Peripheral D Records

Similar to nested D records, some inmates have D records that precede or follow terms in a larger term history. We call these D records peripheral because they either begin a history or conclude it. But like with nested D records, we must resolve peripheral records to create logically consistent term histories. Failure to integrate these records will lead to an undercount of inmate populations like the one described in Section 3.3. Fortunately the imputation process here follows exactly the same process as outlined for nested D records, except that instead of defining the boundaries of the interval using B and A records from other terms, we substitute the boundaries as constrained by the reporting period for states. For peripheral D records that <u>precede</u> a term history, the boundaries of the interval are defined by the first day that records are reported up to the first reported A record. For peripheral D records that <u>follow</u> a term history, the interval is defined by the last B record, up to the end of the reporting period. Using these redefined intervals, we perform the same steps and imputations as with nested D records.

Where we observe every potential peripheral D record on an interval, we treat this as if no time has been spent away from prison. Tables 8 and 9 below illustrate these cases. Table 8 shows peripheral D records that precede the start of a term history, and Table 9 shows D records that follow the end of a term history.

Table 8

1999		2000		2001		2005	
	D(·)		D(·)	A(1)	D(·)		D(·)

Table 9

	2006		2007		2008		2009	
D(·)		D(·)	B(1,2)	D(·)		D(·)		D(·)

For preceding D records (Table 8), we drop the following A record in favor of the admission date in the D records. For following D records (Table 9), we drop the preceding B record and treat the term as ongoing (i.e., blank).

Where we do not observe every potential peripheral D record on an interval, we do not treat the prison stay as continuous and, instead, impute an additional term. Tables 10 and 11 below illustrate these cases. Table 10 shows peripheral D records that precede the start of a term history, and Table 11 shows D records that follow the end of a term history.

Table 10

1999		**2000**		**2001**		**2005**	
	D(·)			A(1)	D(·)		D(·)

Table 11

	2006		2007		2008		2009	
D(·)		D(·)	B(1,2)					D(·)

To impute term admission dates, we identify the admission date of the earliest reported D record on the interval between terms. For preceding D records, we simply use the recorded admission date. For following D records,

1. if the admission date is equal to or earlier than the date of the B record for the first term, we drop that release record.
2. otherwise we introduce a new term that start on the admission date reported in the D record. In the event that no admission date is recorded on the D record, we impute an admission date according to equation [6].

To impute release dates, we identify the latest D record on the interval. From this date, we impute a release date by selecting a date at random that lies between the date we are using and the date of either (a) the following reported set of D records, or (b) the A record of the second term, whichever comes first. This imputation is identical for both types of peripheral records and simply uses equation [7] from above. Also, if the latest D record that we *actually* observe is the last D record we can *possibly* observe, then we leave the term end date as ongoing (i.e., blank).

Finally, if after the imputation process possible D records can be observed between created and observed terms, for either case, then we combine terms by dropping the appropriate records.

4.2 Invisible Terms without D Records

Sometimes terms exist but we cannot see any A, B or D records. Consider the example of a state that reports A, B and D records in 2002, and again in 2004, but does not report any records for 2003. In this scenario, we would not be able to observe those inmates who were both admitted and released in 2003, though such inmates most likely exist. Where we see gaps in reporting, our first step is to contact the state and request the missing data. In the event the state cannot provide those data, we impute these terms. Our overall approach is to sample from other parts of the data to create a reasonable approximation for how the missing data would appear on average. We start with the assumption that terms in the periods immediately before and after the gap are similar in nature to those missing terms. We use the characteristics of these observable terms to describe the nature and frequency of invisible terms within the gap. More explicitly, we take the following steps:

1. Characterize those terms that we cannot identify. In the above example, these are terms that begin and end in a single calendar/reporting year.
2. Identify and isolate two pools of characteristically similar terms. The first pool should use terms for a similarly-sized period immediately before the gap; in this example, this would be terms that begin and end in 2002. The second pool should use terms for a similarly-sized period immediately after the gap; in this example, these are terms that begin and end in 2004.
3. Identify each period (i) that occurs between the two pools. In this example, one period exists ($i = 1$) between the two pools−the year 2003. Denote the total number of periods as P.
4. Assemble new groups of terms for each period (i) by randomly selecting a proportion of terms (ω_i) from pool 1, and a proportion of terms ($1 - \omega_i$) from pool 2, where

$$\omega_i = i/(1 + P), \text{ for each } i \in [1,P].$$

These new groups will augment the existing term-level dataset.

5. Translate the term admission and release dates of the selected groups so that they begin and end in the unobserved period, rather than the period in which they were originally observed. To do this, simply add days (D_{ij}) to each date within terms in period i that were selected from pool j, where,

$$D_{ij} = (\text{Start date of period } i) - (\text{Start date of pool } j)$$

In this context of the example, this amounts to simply adjusting the year of the admission/release.

6. Finally, append these new terms to the final term-level datafile. Again, these new terms are not meant to replace the terms from pools 1 and 2 that are already in the final dataset, they simply augment the data.

5. Final Adjustments to Terms and Term Histories

After the processes and imputations described above have been applied, there may be residual inconsistencies in terms histories. The first problem is that term histories may contain individual (even unambiguous) terms that overlap in time−a new admission occurs before the prior term has ended. The second problem is that terms may indicate that a release happened before an admission−a clearly erroneous result. Such errors happen rarely and, as such, have minimal effect on statistics. Nevertheless we take some final steps to resolve these errors. Where the data indicate that terms overlap, and these terms seem sensible, we combine them into a single, continuous term. Term histories with a releases coinciding with new admissions are counted among these overlapping terms, since it is unlikely that an offender is returned to prison on the day they are released. We also delete terms where releases occur before admissions, and prison terms where the admission and release occur on the same day.

6. Combining Prison and PCCS Terms and Term Histories

The final step in constructing valid offender-based longitudinal histories is to assess whether the implied prison and PCCS terms resulting from the earlier process produce cycles of incarceration and supervision that are logically and statistically valid. For individual offenders, prison and PCCS terms are arranged in chronological order and inspected for error/inconsistency. In some cases, this inspection reveals that, overall, data do not agree well. For example, an unusually large number of prison terms may be embedded within PCCS terms, or vice versa. In those cases, we set the data aside for further investigation with the state and do not combine prison and PCCS term histories into larger longitudinal histories. However in many cases the data area naturally in strong alignment, e.g. PCCS terms begin on the same day as a prison release, though not all terms are without error. In those cases, we further process and combine prison and PCCS term histories into larger longitudinal histories, resolving inconsistencies for those cases not in natural agreement. The purpose of this section is to describe those additional processing steps.

Similar to before, we develop some specific language around this exercise to assist in its description. The existing terminology of **ambiguous** and **unambiguous** appears in this section and is consistent with earlier definitions. We also describe movements from prison to PCCS and vice versa as **transitions,** similar to the use of the word *sequence* from before**.**

Lastly, we use the language **consistent**, **marginal**, and **inconsistent** to describe the logical nature of those transitions. Each of these descriptions is mutually exclusive and comprises the universe of transitions. **Consistent** cases are transitions where the implied timing (i.e., of admission and release dates) is logically sound and require no additional manipulation. **Marginal** cases are transitions where each terms partially overlaps the other. **Inconsistent** cases are transitions where one terms appears entirely within another term. Marginal and inconsistent cases both require adjustment. Depictions of these cases appear in the following sections. Also, a summary guide of transitions and applied resolutions can be viewed in the Quick Reference Guide (in Section 9).

6.1   General Principles/Rules

There are some general principles that have guided the construction of the specific resolutions described in this section that readers should be aware of. The first is that most rules have been designed under the assumption that recording practices for and data elements from PCCS records are largely driven by the administrative needs and practices of states agencies, rather than by a need for strict accounting by NCRP standards. More specifically, dates associated PCCS admission and releases are assumed to be (often) based on administrative actions rather than a true daily accounting of placement onto and out of supervision. For example, it would not be uncommon for a supervision officer to admit a prisoner onto PCCS before that prisoner has physically been released from prison, in order to start a casefile in preparation of an eventual release. Likewise, a supervision officer may wait several days, weeks or even months before formally recording a revocation after an offender has already returned to prison because the supervision officer is awaiting the outcome of prosecutorial decisions. The result is that more flexibility is generally associated with PCCS admission and release dates. Less flexibility is given to prison admission and release dates, which are generally assumed to be a stricter accounting of a person's physical placement, since recording of this kind is more important for corrections agencies. As a result, rules developed around transitions typically defer to the prison admission and release dates as the accurate date of placement when dates disagree.

Another principle of constructing and combining histories is that terms with release dates and admission dates occurring on the same day are excluded. We do this because these terms are not the type of terms NCRP seeks to identify. Prison terms with admission and release on the same day often represent offenders that have served their entire term in a county jail. Administrative records are created in DOC systems for accounting purposes, but these offenders don't spend any nights in the prison. PCCS terms with admission and release on the same day can result when offenders violate parole before they ever leave the prison. For example, if a sex offender doesn't have an allowable residence, they have violated the condition of their parole and they remain in prison. Again administrative records are created for accounting purposes but these offenders are never truly under supervision.

Finally, there are some cases where individuals are released from PCCS, then readmitted extremely quickly. Such gaps are unlikely to be true released from PCCS; instead we treat these gaps the result of as administrative issues, where e.g., an offender could have been transferred to a new supervision officer having never truly left supervision. Specifically, we combine consecutive PCCS terms separated by less than seven days into a single term where the admission date is associated with the term before the gap in service, and the release date is associated with the term after the gap in service

6.2 Resolving Transitions

In this section we describe our steps for resolving marginal and inconsistent transitions in combining prison and PCCS terms. We organize the discussion by the type of transition. In general, there are two types of transitions:

(1) a PCCS admission preceding a prison admission (PCCS → Prison)
(2) a PCCS admission following a prison admission (Prison → PCCS)

These are not mutually exclusive groupings in the sense that an individual term can be both part of case (1) and case (2) simultaneously, depending on the term histories. Also, the nature of the resolutions employed is such that the order of resolution is largely inconsequential, however operationally we resolve all instances of (1) first and instances of (2) second.

6.2.1 PCCS to Prison Transition

Transitions from PCCS to prison appear one of four ways:

Consistent cases: $(E) \rightarrow (F, A) \rightarrow (B)^{18}$
Marginal cases: $(E) \rightarrow (A)^{19} \rightarrow (F) \rightarrow (B)$
  $(E) \rightarrow (F) \rightarrow (A) \rightarrow (B)$
Inconsistent cases: $(E) \rightarrow (A) \rightarrow (B) \rightarrow (F)^{20}$

The notation (F, A) indicates that the PCCS release date and prison admission date occur on the same day. For the **consistent** case, the dates are aligned as expected and no adjustment is needed.

---

[18] The B record may occur outside the reporting window.

[19] If the A record occurs on the same day as the preceding E record, this record would be considered an inconsistent case $((A) \rightarrow (E) \rightarrow (F) \rightarrow (B))$ from earlier.

[20] Both the B and F record here can also occur on the same day, including outside the reporting window.

For **marginal** cases, the reentry into prison movement either occurs before a revocation [(A) → (F)] or after the release from PCCS [(F) → (A)]. For the former, a likely explanation is that an offender returned to prison while still on PCCS but was not formally revoked until sometime later for administrative reasons. Our solution is to align the PCCS release to coincide with the prison admission date. For the latter case, the return to prison after the conclusion of PCCS is not a logically inconsistent transition. However, it may also be true that the PCCS release is intended to represent a revocation, and that slippage in the recorded timing of these events makes them incorrectly appear unrelated. To assess this we (i) examine the time elapsed between PCCS release and prison admission, and (ii) inspect the PCCS release and prison admission codes. In cases where the PCCS release is recorded as a '*Revocation*' or the prison admission is recorded as a '*Revocation*,' we align the PCCS release to coincide with the prison admission date. Otherwise, if the gap between release and admission is less than 30 days, we align the PCCS release to the prison admission date. Finally if none of those conditions hold, we leave the term as-is.

For **inconsistent** cases, an offender serves a complete prison term while continuously under supervision. Such events are relatively inexplicable. In these cases we assume that the prison term is a legitimate prison stay and our solution is two-fold. If the prison release record indicates an unconditional release from prison, we treat the F record as a late accounting of a revocation and move it to the prison admission date such that (E) → (A) → (B) → (F), becomes E→ (F,A) → B. Otherwise we impute a revocation event such that the transition (E) → (A) → (B) → (F), becomes E→ (F,A) → (B,E) → F.

Rules used to resolve inconsistencies in PCCS → prison transitions *ignore* whether terms are derived through ambiguous or unambiguous record sequences.

6.2.2   Prison to PCCS Transition

Transitions from prison to PCCS appear one of four ways:

Consistent cases: (A) → (B, E) → (F)[21]
Marginal cases: (A) → (E)[22] → (B) → (F)
                          (A) → (B) → (E) → (F)
Inconsistent cases: (A) → (E) → (F) → (B)[23]

The notation (B, E) indicates that the prison release date and PCCS admission date occur on the same day. For the **consistent** case, the dates are aligned as expected and no adjustment is needed.

---

[21] The F record may occur outside the reporting window.

[22] This E record may occur on the same day as the preceding A record.

[23] Both the B and F record here can also occur on the same day, including outside the reporting window.

Unlike the rules described in previous description, rules to resolve inconsistencies in prison → PCCS transitions depend whether the <u>prison term</u> was derived through an ***unambiguous or ambiguous*** record sequence. We described both sets of rules here.

Transitions with Unambiguous Prison Terms

For **marginal** cases, the placement onto PCCS either occurs after a release from prison but not immediately [(B) → (E)] or before a release from prison [(E) → (B)]. For the former, the cause of the gap in service is unknown, but a possible explanation may be that an offender moved to another jurisdiction (e.g., to stand trial out of state) then later returned to the jurisdiction of the original state. On the other hand, the gap in service may simply reflect idiosyncratic (administrative) recording practices. Given this uncertainty, our solution is to align the PCCS admission to coincide with the prison release date *if the PCCS admission occurs within 30 days of the prison release*. Otherwise, we leave the dates as they appear. For the latter case, the admission to PCCS prior to release from prison likely results from a supervision officer opening a PCCS record in anticipation of an upcoming release. Alternatively, this may reflect a transition to halfway facilities, where an offender is still in DOC custody, but where they also have a PCCS admission in anticipation of eventual release. To resolve these cases, we align the PCCS admission to coincide with the prison release date.

For **inconsistent** cases, an offender serves a complete PCCS term while continuously in prison. Again, such events are relatively inexplicable. Our default position is to defer to the prison term as correct and drop the PCCS term, unless it can otherwise be shown to be accurate. To judge the accuracy of the PCCS term, we inspect the type of admission and release. If the PCCS admission is recorded as a 'Release from Prison' and the release is recorded as a 'Revocation', we impute a release from and return to prison so that (A) → (E) → (F) → (B), becomes A → (B,E) → (F,A) → B. Otherwise, we drop the PCCS term so that (A) → (E) → (F) → (B) becomes A → B.

Transitions with Ambiguous Prison Terms

The manner in which ambiguous terms are formed dictates how we resolve inconsistencies in transitions. In some cases, ambiguous recording resulted in **dropping records**. For example, when observing an A(1)➔A(2)➔B(1,3) prison term, our decision rules would drop the A(2) record and create an A(1)➔B(1,3) term. In other instances, ambiguous recording resulted in **imputing records.** For example, if we observe an A record with no subsequent B record and an inconsistent accounting of D records, we impute a release date randomly between the last observed D record and the date of the first unobserved D record (1 year later). Finally, some ambiguous terms were formed by **combining terms**. We look at the cause of the ambiguity as well as the length of time between records in determining which action to take.

For **marginal** cases where the PCCS admission occurs before a release from prison [(E) → (B)], we assume that it is a variation of the consistent case and that the ambiguity has caused the distortion. If the resolution of the ambiguous case involved **imputing** records, we align the prison release date to the PCCS admission date, except in cases where the PCCS term is ambiguous and occurs in a year other than the prison release. In that case we align the prison release date to the PCCS admission date. We do this under the assumption that we were more likely to have chosen the incorrect prison date in the case of an imputed record. If the resolution of the ambiguous case involved **dropping** records or **combining** terms, we always align the PCCS admission date to the prison release date.

For **marginal** cases where the PCCS admission occurs after a release from prison [(B) → (E)], the ambiguity of the prison term is ineffectual if ambiguity resulted from **dropping** records or **combining** terms. In those cases, we continue the convention for unambiguous versions of this transition, i.e., we align the PCCS admission to coincide with the prison release date if the PCCS admission occurs within 30 days of the prison release. Otherwise, we leave the dates as they appear. If the ambiguity of the prison term involved **imputing** records, we use the type of admission to PCCS and type of release from prison to determine if the transition is implied to coincide. If the release from prison indicates a supervised release and the admission to PCCS indicates a release from prison, then we align the prison release to coincide with the PCCS admission. Otherwise, we leave the dates as they appear.

For **inconsistent** cases where an offender serves a complete PCCS term while continuously under supervision, the ambiguity of the prison term may be driving the ambiguity of the transition. For cases where the ambiguity of the prison term resulted from **dropping** records or combining overlapping terms, we defer to the PCCS term as correct and impute a release from and return to prison so that (A) → (E) → (F) → (B), becomes A → (B,E) → (F,A) → B. If the ambiguity of the prison term resulted from **imputing** records, inspect the type of PCCS admission and release. If the PCCS admission is recorded as a 'Release from Prison' and the release is recorded as a 'Revocation', we impute a release from and return to prison so that (A) → (E) → (F) → (B), becomes A → (B,E) → (F,A) → B. Otherwise, we drop the PCCS term so that (A) → (E) → (F) → (B) becomes A → B.[24]

7. Verifying Results − Replicating Prison Stocks

Our last step prior to finalization of the term-level dataset is to assess the validity of the data by comparing prison stocks implied by the constructed term records with prison stocks reported on the D records. In effect, we use the constructed term records to generate an estimate of prison stocks. We then compare this estimate of prison stocks with the actual prison stocks reported on the D records. Because offenders are sometimes temporarily absent

---

[24] Appendix 10 notes an exception to imputation rules made for Kentucky.

from prison when the state assembles D records, the comparison is never exact but it should be within a few percentage points.

Conducting this comparison enables two essential insights into the validity of the constructed terms. First, we the constructed terms may significantly over or under reports prison stocks. Where there is considerable variance, we investigate collaboratively with the state about the data collection and reporting process for NCRP to identify sources of this variance. The algorithm described in this paper is partly a result of developing rules to make the flow of prison admissions and released conform with the stock. Sometimes the iterative procedure caused us to diagnose and correct problems with A, B or D records. In this regard the algorithm described in this paper was empirically informed.

Second, we verify that the estimated prison stocks from the imputed terms follow similar trends over time to that of the reported D records. This helps ensure that, while population estimates may differ by amount, the overall prison population trends mirror each other. Should the stock estimates from the time-series suggest a trend significantly different from the stock trends of the reported D records, we reassess the term record assembly and investigate systemic differences in the data.

8. Technical Notes

Variable information is taken from A, B and D records in the following manner. First, all information is taken from the preprocessed B records where release dates and inmate identifiers in the preprocessed B records (after we change IDs, drop B's with no release dates, etc.) match the release dates and inmate identifiers in the term records. There should be a match in every case where a B record was not imputed. Second, variable information still missing is updated from the A records where admission dates and inmate identifiers in the preprocessed A records (after we change IDs, drop A's with no admission dates, etc.) match admission dates and inmate identifiers in the term records. Again, there should be a match in every case where an A record was not imputed. Together, these first two steps effective map the vast majority of variable from the raw records to the terms in the analysis file.

Third, variable information that is still missing in the analysis file is updated from the raw D records where admission dates and inmate identifiers in the preprocessed D records match admission dates and inmate identifiers in the term records. Only records where no previous match to an A or B record was found are affected by this additional step. Where we find multiple D records that share a common admission date and that also match an admission date for a term record, we take the variable information from the first D record. Lastly, any information that is still missing is updated using variable information from the first observable D record after the start of a term for an individual. If variable information is still

missing, we chronologically cycle though the remaining D records associated with the prison term, replacing missing values when non-missing information exists.

9. Quick Reference Guide

Term Construction Rules

Term or Case	Rule imposed	Imputation	Proportion of all terms reported from A & B records
Unambiguous	None		88 – 98%
Ambiguous			2 – 12%
A→B→B	Drop middle record		< 10%
A→A→B	Drop middle record		< 5%
Overlapping Terms			< 2%
Unambiguous	Combine Terms	Start = min(A), End = max(B)	< 1%
Ambiguous	Impose refinements	A(1) → B(1,2)…A(3) …B(1,4) becomes, A(1) → B(1,2)…A(3) → B(3,4), where an imputation is done for A(3)…B(1,4) A(1)…A(2)…B(1,3) becomes A(1) → B(1,3) where A(2) is dropped	< 2%
Admission Observed Last	Impute release if inmate is absent in final stock	$d_B = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$ if any D exists, and Release = (Admission + 30) if no D exists	< 1% with imputations

Term or Case	Rule imposed	Imputation	Proportion of all terms reported from A & B records
Release Observed First	Impute admission if inmate is absent in initial stock	$d_A = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$ if any D exists, and Admission = (Release − 30) if no D exists	< 1% with imputations
Partially Unreported Terms			0% in 95% of states
A with no B	Impute release between the last observed D and the first missing D	$d_B = (d_{D1} + 1) + (r)(d_{D2} - d_{D1})$	0% in 95% of states
B with no A	None	Use admission date from B record	0% in 95% of states
Invisible Terms			< 2% - 3%
Only D Records	Repeat steps from rows 8 and 9	See rows 8 and 9 in this table.	< 2%
Nested			< 1%
Continuous	Combine Terms	Drop the preceding release and following admission dates	< 1%
Not continuous	Impute admission and release dates	Impute an admission date using the date from D record. Impute a release record as in row 11.	< 1%
Peripheral			< 1%

Term or Case	Rule imposed	Imputation	Proportion of all terms reported from A & B records
Continuous	Combine Terms and impute a new admission/release date	Either drop the preceding release and impute a new release as in row 11, or drop the following admission and impute a new admission using the date from D record.	< 1%
Not continuous	Repeat steps from row 17	See row 17 in this table.	< 1%
No D records (i.e., where gaps in reporting exist)	Impute terms through a sampling process	See Section 4.2	0% in 95% of states

Transition Rules

General Rules			
E-F terms of less than 5 days are deleted.			
Gaps between consecutive EF of less than 5 days are always combined.			
Transition	**Sequence**	**Rules**	**Result**
Prison → PCCS with unambiguous prison terms	A → E → B → F	Align E to B.	A → [B, E] → F
	A → B → E → F	For E <= 30 days of B, align E to B.	A → [B, E] → F
		For E > 30 days after B, no adjustment.	A → B → E → F

	A → E → F → B	if E code ≠ *'release from prison'* or F code ≠ *'revocation'*, then drop E and F.	A → B
		if E code = *'release from prison'* and F code = *'revocation'*, then impute B and A.	A → [B, E] → [F,A] → B
Prison → PCCS with ambiguous prison terms	A → E → B → F	If we imputed a record and PCCS is unambiguous, then align B to E.	A → [B, E] → F
		If we imputed a record, the PCCS term is ambiguous and E occurs in the same year as B, then align B to E.	A → [B, E] → F
		If we imputed a record, the PCCS term is ambiguous and E occurs in a different year from B, then align E to B.	A → [B, E] → F
		If we dropped a record, then align E to B.	A → [B, E] → F
		If we combined terms, then align E to B.	A → [B, E] → F
	A → B → E → F	If we dropped a record or combined terms, for E <= 30 days of B, align E to B.	A → [B, E] → F
		If we dropped a record or combined terms, for E > 30 days after B, no adjustment.	A → B → E → F
		If we imputed a record, the B code = *'release to supervision'* and E code = *'release from prison'*, then align B to E	A → [B, E] → F
		If we imputed a record and either the B code ≠ *'release to supervision'* or E code ≠ *'release from prison'*, then no adjustment	A → B → E → F
	A → E → F → B	If we dropped a or combined terms record, then impute B and A.	A → [B, E] → [F,A] → B

		If we imputed a record, the E code = *'release from prison'* and F code = *'revocation',* then impute B and A.	A → [B, E] → [F,A] → B
		If we imputed a record or combined terms, and either the E code ≠ *'release from prison',* and F code ≠ *'revocation',* then drop the E and F.	A → B
PCCS → Prison	E → A → F → B	Align F to A.	E → [F, A] → B
	E → F → A → B	if F code = *'revocation'* or A code = *'revocation',* align the F date to the A date.	E → [F, A] → B
		If neither F nor A codes are revocations and A <= 30 days after F, align the F date to the A date.	E → [F, A] → B
		If neither F nor A codes are revocations and A > 180 days after F, no adjustment.	E → F → A → B
	E → A → B → F	If the B code indicates an unconditional release, then move the F date to the A date.	E → [F, A] → B
		If the B code indicates any other than an unconditional release, treat as a revocation and readmission to PCCS. Impute F and E.	E → [F, A] → [B,E] → F
PCCS → PCCS	E → F → E → F	If average F - E < 30 days, then drop middle F and middle E.	E → F
		If average F - E >= 30 days, no adjustment.	E → F → E → F

10. Appendix

As noted in Section 2.3, we modify our approach for handling ambiguous cases in 8 states: Alaska, Indiana, Kentucky, Maine, Maryland, Nebraska, Pennsylvania and West Virginia. In these states, detailed analysis of term records, raw data records and direct discussions with State DOCs (in Maryland and West Virginia) has led us to conclude that observed ambiguous terms are the product of recidivistic events. Thus for the final term records to accurately reflect the reality in these States, we modify this portion of the process. The following is a description of this modification.

The process of resolving ambiguous cases starts with the estimation of a statistic *P*. *P* represents the average probability that an inmate will be recorded in any given D record over a specified interval. It is calculated as the ratio of D records actually observed to D records possibly observed, over the interval. A different value for *P* is calculated for each of the five ambiguous cases described earlier in Table 4. For ambiguous cases with consecutive admissions ("A→A→B" as a shorthand notation), *P* is calculated based on the interval between admissions. For ambiguous cases with multiple release records sharing admission dates ("A→B→B" as a shorthand notation), *P* is calculated on the interval between releases. Using this statistic *P*, impute the associated (missing) admission or release for each ambiguous case.

In the case of A→A→B, we impute a release record, so that we are left with two distinct terms, A→(B) & A→B. We impute the date of this release record as follows: Let d_{A1} be the admission date for the first A record and let d_{A2} be the admission date for the second A record. Then the release date for the imputed B_1 record is,

$$[1] \quad d_{B1} = d_{A1} + (P)(d_{A2} - d_{A1})$$

In the case of A→B→B, we impute an admission record, so that we are left with two distinct terms, A→B & (A)→B. We impute the date of this admission record as follows: Let d_{B1} be the release date for the first B record and let d_{B2} be the release date for the second B record. Then the admission date for the imputed A_2 record is,

$$[2] \quad d_{A2} = d_{B1} + (1-P)(d_{B2} - d_{B1})$$

Equations [1] and [2] above depict the general formulas for imputation, however, in practice we compute (and use) separate values of *P* (i.e., $P_1, P_{2, ...,} P_j$) for subgroups of inmates within an ambiguous case where these subgroups are defined according to the number of potential D records spanned along the interval over which *P* is calculated. Specifically, P_1 is calculated for inmates whose interval spans exactly 1 D record, P_2 for those whose interval spans exactly 2 D records, P_3 for those whose interval spans exactly 3 D records,

etc. There are also inmates whose interval does not overlap any D records. For this group no calculation (P_0) can be performed since the denominator of P for this subgroup must be zero. As a result, we use P_1 in performing imputations for this subgroup.

The justification for this imputation procedure may seem obscure, however we emphasize that a sensible process is needed in order to avoid making gross errors.