**7 November 2012**

Master Person Index

for the

MHS Data Repository (MDR)

(Version 1.01.00)

Current Specification

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date** | **Para/Tbl/Fig** | **Originator** | **Description of Change** |
| 1.00.00 | 07/20/2006 | * Whole document |  | * Initial versioning. |
| 1.01.01 | 11/07/2012 | * Whole document | Joe Shoemaker with Wendy Funk’s approval | * Person ID and PARC updates |

MDR Encounter Data are not always fully identified with patient attributes. This specification describes the use of the MDR Master Person Index (MPI) to improve the patient information in that encounter data.

The MPI Merge File:

The MPI is generated monthly by the MDR’s processing of the DMDC VSAM extract sent to EI/DS. This specification assumes the VM6 formatted feed is used to generate an MPI with the fields shown below in Table 1; the VM6 functional specification describes the actual creation of the file.

**Table 1: MPI Fields**

Person Identifier (EDI-PN)

Person Social Security Number (PSSN)

Sponsor Social Security Number (SSSN)

Legacy DEERS Dependent Suffix (DDS)

Person Association Reason Code (PARC)

Relationship[[1]](#footnote-1) (1-4 for dependent child, sponsor, spouse, and other, respectively)

Member Relationship Code

Gender

Date-of-Birth

Last Name

First Name

Because the MPI is generated anew every month and no records are purged from past months in the source, there is never a need to retain or use an older MPI. At the time of the monthly snapshot, every relationship known to DEERS since 1 January 2000 has a record in the MPI. One individual may have multiple MPI records because of multiple sponsors associated with him/her; however, one part of MPI preparation is to “dedup” the MPI to remove any instances where more than one MPI record match the first three fields listed above.

Encounter Data:

All encounter data types are candidates for merging to the MPI to improve the patient identification, based on one or more fields within the encounter data providing partial patient identification. Table 2 below summarizes the encounter data types and the type of identifying information within each.

All records from each encounter type that have neither been successfully merged to the MPI in the past nor clearly identified as records that will never be in the MPI are retried each month when a new MPI becomes available. It is acceptable to keep retrying even the records that will never be in the MPI.

**Table 2: Person Identifiers by Data Type (in Raw Record before MPI Merge)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Type** | **EDI-PN** | **Sponsor SSN** | **DDS** | **Relationship** | **Gender** | **DOB** | **Name** |
| Appointment | U | U | S | A2 | A | A | A8 |
| SADR | U | U | S | A2 | A | A | N |
| SIDR | U | U | S | A2 | A | A | A[[2]](#footnote-2) |
| DC Ancillary | U | U | N | A[[3]](#footnote-3) | N | N[[4]](#footnote-4) | N |
| PDTS | U | S | S | N | A | A | A8 |
| HCSR | N | U | U | A[[5]](#footnote-5) | A | A | A[[6]](#footnote-6) |
| ATOH | U | U | N | A[[7]](#footnote-7) | A | A | A[[8]](#footnote-8) |
| TED | U | U | N | A7 | A | A | A7 |

A: Always present

U: Usually, but not always, present

S: Sometimes present

N: Never present

Preparation

Prior to the MPI merge, each encounter data type stores the demographic data it received in “raw” fields. Merge fields matching the MPI fields are then prepared. If the field is already populated with a “valid value” (see Table 3), the valid value is copied to the merge field of corresponding name. If the value is empty or not valid, no value is copied to the merge field.

PDTS is unique in that its raw “Subscriber ID” field may be either an EDI-PN or an SSSN and DDS. While the Subscriber ID (raw) field is not altered, the absence of any other family identifiers means that there is only one field each for EDI-PN, Sponsor SSN, and DDS. Whichever of these came in the Subscriber ID field is written from it, and the remaining field(s) populated from the MPI merge process.

**Table 3: Valid Values for MPI Merge Fields**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Valid values** | **Comment** |
| EDI-PN | Exactly 10 characters in length, all of which are numeric digits, and the first of which is a “1” |  |
| Person SSN  Sponsor SSN | Exactly 9 characters in length, all of which are numeric digits, excluding the values “000000000” and “999999999” |  |
| DEERS Dependent Suffix | Exactly two characeters in length, allowed values are 01-20, 30-69, and 98 |  |
| PARC | Not blank |  |
| Relationship | “1”, “2”, “3”, and “4” | See derivation column in Table 4 |
| Gender | “F” and “M” | Any “U” or “Z” in VM4 is left empty in MPI |
| Date-of-Birth | SAS date greater than Jan 1, 1870 but less than or equal to begin-date-of care or service date |  |
| Last Name | No numerics or special characters. | Left trimmed of any blanks |
| First Name | No numerics or special characters. | Left trimmed of any blanks |

**Table 4: Deriving Relationship for MPI Merge**

|  |  |  |
| --- | --- | --- |
| **Data Type** | **Raw Field** | **Derivation** |
| Appointment | FMP | If FMP is 01-19 then Relationship = “1”  If FMP is 20 then Relationship = “2”  If FMP is 30-39 then Relationship = “3”  If FMP is in list 40, 45, 50, 55, or is within 60-69, then Relationship is “4”  ELSE Relationship is “2”. |
| SADR | FMP |
| SIDR | FMP |
| DC Ancillary | FMP |
| PDTS | DDS (if populated) | If DDS is 01-19 then Relationship = “1”  If DDS is 20 then Relationship = “2”  If DDS is 30-39 then Relationship = “3”  If DDS is 40-69, then Relationship is “4”  ELSE Relationship is “2”  . |
| HCSR | DDS |
| ATOH | Member Relationship Code | If MRC inlist (A,) then Rel = “2”  If MRC inlist (B,G,H, I,J,K), then Rel = “3”  If MRC inlist (C,D,E,L) then Rel = “1”  If MRC = “F” then Rel = “4”  Else:  if Age < 18 years then Rel = “1”.  if Age >= 18 and gender is “F” then Rel = “3”.  Else Rel = “2”. |
| TED | Member Relationship Code |

MPI Merge Procedures

Each observation of encounter data is qualified according to the following table, and the selected business rules are applied as described further below. The “?” character is used to indicate that at least one of the fields on the row is desired and invalid.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case** | **Valid EDI-PN?** | **Valid Person SSN?** | **Valid Spon SSN?** | **Valid DDS?** | **Valid PARC?** | **Business Rule** |
| 1 | Y | Y | Y | Y | Y | No Merge[[9]](#footnote-9) |
| 2 | Y | ? | Y | ? | ? | A. MPI-PID |
| 2a | N | Y | Y | ? | ? | A. MPI-PID |
| 3 | Y | ? | N | ? | ? | B. Approx MPI-PID |
| 3a | N | Y | N | ? | ? | B. Approx MPI-PID |
| 4 | N | N | Y | Y | ? | C. MPI-SSN-DDS |
| 5 | N | N | Y | N | ? | D. Approx MPI-SSN |
| 7 | N | N | N | n/a | n/a | No Merge |

A. MPI-PID

The basic MPI-PID merge joins the encounter record to the MPI by either EDI-PN or PSSN. Upon success, the derived DDS, PARC, and, if invalid, whichever ID was not used for the join (i.e., PSSN or EDI-PN) are set to the values of the respective fields from the MPI. Upon failure, the derived DDS and PARC are set to blank.

* The join succeeds if the result is a single MPI record that has the same sponsor SSN as the encounter record.
* The join succeeds if the result is a single MPI record and that record does not have the same sponsor SSN as the encounter record. The derived sponsor SSN should be set to the value of the sponsor SSN from the MPI.
* The join succeeds if the result is multiple MPI records that all have differing values of the sponsor SSN relative to the encounter record. MPI values are selected from the first matching record.
* The join fails if no MPI records were matched.

If the join fails on EDI-PN and the PSSN is valid, this process is repeated using the PSSN as the key to the join.

B. Approximate MPI-PID

The approximate MPI- PID merge joins the encounter record to the MPI by either EDI-PN or PSSN, used when no valid Sponsor SSN is available. Upon success, the derived DDS, PARC, SSSN, and, if invalid, whichever ID was not used for the join (i.e., PSSN or EDI-PN) are set to the values of the respective fields from the MPI. Upon failure, the derived DDS, PARC, and SSSN are set to blank.

* The join fails if unable to locate a matching record.
* The join succeeds if one or more matching records are found. To determine which record to use for values, the following four tests, in order, are applied to each matching record by comparing it to the encounter record. If any test is passed by a single MPI record, that MPI record is selected and the process ends at that test.

1. Gender and date-of-birth are exact matches.
2. Date-of-birth is an exact match.
3. Gender is an exact match, and dates-of-birth are within two months and one day of each other.
4. Dates-of-birth are within two months and one day of each other.

If no MPI records pass any of the preceding tests, the first record that matched the join key, either EDI-PN or PSSN, is selected, and any further testing is abandoned.

If a single MPI record was not selected from the previous tests, then the multiple MPI records that passed the highest ranked test are filtered by the closeness of their names[[10]](#footnote-10). First, only the MPI records that have the closest last names are retained. Second, of those retained, only the MPI records that have the closest first names are given any further consideration. The first MPI record in the final surviving set is selected.

C. MPI-SSN-DDS

The basic MPI-SSN-DDS merge joins the encounter record to the MPI by the sponsor SSN and DDS, and is used when no valid EDI-PN or PSSN is present.

* If the join succeeds, then the derived EDI-PN and PSSN fields are set to the values of the EDI-PN and PSSN from the MPI.
* If records matching SSN but not DDS are found, the process below is followed.

D. Approximate MPI-SSN

The approximate MPI-SSN merge joins the encounter record to the MPI by the sponsor SSN and relationship, and is used when no valid EDI-PN or PSSN is present and when either the DDS is invalid or not found in an initial MPI merge attempt.

* If the join fails to find a record, then the derived EDI-PN and PSSN fields are set to blank, and the derived DDS and PARC fields, if not valid, are set to blank.
* If the join succeeds, then the following four tests, in order, are applied to each matching record by comparing it to the encounter record. If any test is passed by a single MPI record, that MPI record is selected and the process ends at that test.

1. Gender and date-of-birth are exact matches.
2. Date-of-birth is an exact match.
3. Gender is an exact match, and dates-of-birth are within the DOB window.
4. Dates-of-birth are within the DOB window.

If no MPI records pass any of the preceding tests, then the derived EDI-PN and PSSN are set to blank, and the derived DDS and PARC fields, if not valid, are set to blank. If some MPI records passed the previous tests but a single MPI record was not selected, then the multiple MPI records that passed the highest ranked test are filtered by the closeness of their names. First, only the MPI records that have the closest last names are retained. Second, of those retained, only the MPI records that have the closest first names are given any further consideration. The first MPI record in the final surviving set is selected.

If a single MPI record is selected, the derived EDI-PN, PSSN, DDS, and PARC fields are assigned the respective values from that MPI record.

Implementation of the MPI Merge Results

If “no merge” is the business rule, the encounter record need not use the MPI Merge process, and the derived fields are determined from the raw fields (Case 1) or are made blank (case 7)

Derived fields are kept as obtained from the MPI merge, unless the merge was unsuccessful. If the other fields in the encounter record infer the patient will never be recorded in DEERS as a DEERS eligible, the derived EDI-PN will be set to “NOTINDEERS”, the DDS to “20”, and the Relationship to “2”.

1. If the record has blank or null for member relationship code, assign it as a sponsor relationship (“2”). [↑](#footnote-ref-1)
2. 14 character, format unclear [↑](#footnote-ref-2)
3. Patient Category Code, Family Member Prefix both infer relationship. [↑](#footnote-ref-3)
4. Can use “Age” and “Date of Service” for rough approximation of birth year. [↑](#footnote-ref-4)
5. “Common Bencat”, where 2 and 4 are sponsors, 1 and 3 dependents or possibly unknown. [↑](#footnote-ref-5)
6. Last name, a comma, first name [↑](#footnote-ref-6)
7. Member relationship code, where B, G, H, I, J and K are spouses; C, D, E, and L are dependent children; F are other relationships; and A, Z, and any other value are treated as sponsors. [↑](#footnote-ref-7)
8. Separate fields for last name, first name, (TEDs also have fields for middle name, and cadency, PDTS has a field for middle initial.) [↑](#footnote-ref-8)
9. This describes the current process, which fails to resolve records where the EDI-PN and Sponsor SSN/DDS refer to different people. An enhancement to the MPI merge could verify these pointers represent the same person, and when they do not, pick a most likely patient identity based on gender and birthdate. [↑](#footnote-ref-9)
10. The SAS utility SPEDIS is an acceptable means to determine “closeness” of names. [↑](#footnote-ref-10)