1.0 Introduction
It is well established that Patient Identity (PI) Integrity impacts the success and effectiveness of the clinical and financial operations of an organization. The Key Performance Indicators (KPI) presented here focus on PI Integrity. The focus of this document is a limited set of performance measures considered to be key indicators for patient identity management and governance. They are provided as tools for defining and measuring the quality of activities related to an organization’s effectiveness at maintaining a high state of PI Integrity. These KPIs quantify the state of existing performance of business process and the technology that supports the Master Person (Patient) Index (MPI) including algorithms, matching logic, and threshold selections, among others. These KPIs also provide management tools to identify business process variances across time, location, source, and employee. Using these indicators, an organization can quickly identify where to focus remediation, quality improvement, business re-alignment activities and resources. When selecting an MPI system, an organization should require that the system have functional capabilities to produce standard automated reports that provide these basic data and performance indicators as an ongoing management tool. It is anticipated that these KPIs will grow in future iterations as industry gains practical experience and vendors build systems with more sophisticated data collection capabilities. A Glossary of Terms can be found at the end of this document.

Information governance or data governance relates to these KPIs, as these measures may be a good indicator of the “health” of data to support strategic, business, and clinical objectives. In designing and executing data governance, organization will need to decide which domains of data to master, with the patient domain being a key. Organizations may use this patient domain data to create relationships or hierarchies of data, or establish analytics for clinical and business initiatives. Thus the KPIs become even more important as data governance and analytics becomes mainstream.

It should also be noted that many organizations will store person data, beyond the definition of a patient, in the patient domain of the Master Data Management (MDM) software, or the Enterprise Master Person (Patient) Index (EMPI). For instance, employees, guarantors, subscribers, and members might have data within the MPI.

2.0 What Data Should be Collected to Measure Performance?
To measure performance, it is important to determine in advance the data needed for collection. The following data elements are considered key to measuring performance in patient identification processes within an Enterprise Master Person (Patient) Index (EMPI) setting. These are the data that comprise the
numerator and denominators for the various formulas used to report and measure performance. Some of these data should be generated by the system or algorithm while others must be collected manually. Depending on the design and aggressiveness of the MPI management process, these data may be collected concurrently or retrospectively during the normal course of business and may be used for EMPI management on a daily, weekly, monthly, and yearly basis.

2.1 EMPI Database Size (EDS): The total number of records or lines stored in the database. This includes individual categories such as unique individuals, duplicates, personnel listings for security access control, quality control names, pseudo names, test names, un-reconciled records, and abandoned registrations, among others.

2.2 EMPI Person Population (EPP): The number of unduplicated individuals contained in an EMPI system or database. This includes all unduplicated individuals counted in the EDS (see Item 1 above) after cleanup and remediation.

2.3 EMPI Database Growth (EDG): The gross number of new unduplicated records stored in a given time interval (e.g. weekly, monthly) regardless of category.

2.4 EMPI Patient/Person\(^1\) Additions (EPA): The number of new patients who are added to the EMPI for the very first time for a given time frame. All other categories of individuals are excluded.

2.5 Total EMPI Matches (TEM): The total number of potential pair candidates presented by the EMPI algorithm for a given period of time. A computer generated duplicate report will identify the number of candidate pairs for matches based on the algorithm rules or thresholds.

- The total number of individual duplicate patient records used in computing the Duplicate Creation Rate is obtained by dividing the Total EMPI Matches (TEM) by two (2).

2.6 Total EMPI Non-matches (TENM): Total number of true matches missed by the algorithm and identified during normal operation processes that require additional validation of records. This is a measure of algorithm performance.

- Note: There is no easy way to capture this but this is a key metric that reflects the accuracy of the algorithm performance.

2.7 EMPI Database Duplicates (EDD): The numeric count of records that are potential duplicates within the database. The EDD is calculated by subtracting the EPP (unduplicated person count in the database) from the EDS (total lines in database).

\(^1\) In this document we are using the terms patient and person as being synonymous.
2.8 False Positive Matched Pairs (FPMP): The number of candidate pairs generated by the algorithm that are found, after manual validation, not to be matched pairs. These are sometimes referred to as False Positives or False Matched Pairs. It is the number of incorrect matched pair determinations made by the algorithm in a given period of time.

2.9 True Matched Pairs (TMP): The number of pairs generated by the algorithm as well as external business processes that are found, after manual validation, to be confirmed as matched pairs. Sometimes referred to as Adjusted Matched Pairs (AMP).

2.10 False Negative 'non-match' Pairs (FNMP): The number of incorrect EMPI non-match decisions made in a given period of time. The number of candidate pairs identified by the algorithm thresholds to be non-matches but after manual validation are determined to be matched pairs. These are most often found through normal business processes such as patient, physician, or scheduler report. These are not identified by the algorithm. In other words, this is the number of incorrect EMPI non-match decisions made. This will result in an otherwise unidentified duplicate remaining in the database. This is sometimes called a false negative count or pair.

2.11 Indeterminate Matched Pairs (IMP): After manual review, the number of algorithm candidate pairs whose identities could not be validated as being the same individual are called an indeterminate match decision.

2.12 EMPI Matching Validation (EMV): An analysis of the Total EMPI Matches (TEM - see computer generated duplicate report in 2.5 above) which identifies the following: total number of true match pairs, true non-match pairs, false positive match pairs, false negative match pairs and indeterminate match pairs: \( EMV = TMP + TNMP + FPMP + FNMP + IMP \). The EMV total number should always equal the TEM.

2.13 Database Error Growth (DEG): The number of erroneous records added to the database over a given period of time. For example, an abandoned registration saved to the EMPI database.

2.14 Total Registrations Performed (TRP): The total number of registration activities (registration, preregistration, scheduling, office visits, among others) performed in a given time period. These activities may vary by organization depending on the registration and scheduling solutions employed as well as their business model.

3.0 Key Performance Indicators
Most of the above data elements or numeric counts (see Section 2) are used to compute the rates in the following KPI formulas.

3.1 EMPI Database Activity Rate (EDAR) (percent): This rate provides the relative activity of the database. It provides the ratio of the total number of EMPI algorithm matching evaluations (TEM)
performed in a given timeframe to the overall size of the EMPI database (EDS). The formula for calculating this rate is:

\[
\text{EDAR} = \frac{\text{Total EMPI Matches (TEM)}}{\text{EMPI Database Size (EDS)}} \times 100
\]

### 3.2 EMPI Database Duplicate Rate (EDDR) (percent):
This is the percent of paired records in the database that are potential duplicates or multiples. This measure reports a duplicate rate prior to research and validation of the records paired by the algorithm. It is commonly referred to as the Database Duplicate Rate, Duplicate Percentage or sometimes Pair Rate. The formula for calculating this rate is:

\[
\text{EDDR} = \frac{\text{EMPI Database Duplicates (EDD)}}{\text{EMPI Database Size (EDS)}} \times 100
\]

**Example:**
Database line count = 1,000,000 lines; Algorithm identifies 125,000 potential pairs (250,000 records).
Therefore:
\[
\text{EDDR} = \frac{125,000 \times 100}{1,000,000} = \text{a duplicate rate of 12.5%}
\]

### 3.3 Duplicate Creation Rate (DCR)^2:
This is the ratio of newly created duplicate records (numerator) to the opportunity to create a duplicate through various encounters with patients (denominator) in a given period of time. Opportunities to create duplicates would include scheduling events, registration, preregistration, office visits, among others. Denominators may vary by organization depending on the registration and scheduling solutions employed. The total number of individual duplicate patient records is obtained by dividing the Total EMPI Matches (TEM) by 2.

\[
\text{DCR} = \frac{\text{Total number of individual duplicate patient records}}{\text{Total Registrations Performed (TRP)}} \times 100
\]

**Example:**
50 Duplicates Records created in month \( \times \) 100 = 0.5% creation rate
10,000 Registrations Performed

### 3.4 True Match Rate (TMR):
The True Match Rate is the ratio of the number of true match pairs (TMP), as determined after manual validation, to the total number of EMPI matching evaluations (TEM) presented by the algorithm plus those identified by other business processes. The formula for computing the TMR is the number of True Match Pairs (TMP) divided by the total number of potential pair candidates (TEM). This figure provides information on the effectiveness of the algorithm in making matches. A low true match rate may indicate that the programmed matching thresholds may need to be fine-tuned or adjusted.

\[
\text{TMR} = \frac{\text{True Matched Pairs (TMP)}}{\text{Total EMPI Matches (TEM)}} \times 100
\]

---


© December 2012 Healthcare Information and Management Systems Society
3.5 False Positive Match Rate (FPMR): Sometimes referred to as False Match Rate or False Positive Rate. The percentage of incorrectly matched candidate pairs over a given period. This measures the percentage of invalid pairs that have been incorrectly paired by the algorithm. The False Match Rate is the incidence of False Matches made by the algorithm that have been confirmed or validated as not being the same individual. It is computed by dividing the number of False Positive Match Pairs (FPMP) by the total number of potential pair candidates (TEM). This figure provides information on the effectiveness of the algorithm in making matches. A high false match rate may indicate that the programmed matching thresholds are too permissive in their matching criteria and may need to be fine-tuned or adjusted.

\[ FPMR = \frac{\text{False Positive Matched Pairs (FPMP)}}{\text{Total EMPI Matches (TEM)}} \times 100 \]

3.6 False Negative (NonMatch) Rate (FNR): This reports the percent of incorrect EMPI Non-match decisions made in a given time frame. It is the percentage of candidate pairs who should have been matched but were not. These pairs were discovered during the course of business over a given period of time and were not identified by the algorithm. This measure reflects the matching status or decision after review and validation. This is a manual calculation. It must be pointed out that the FNR result will most probably reflect an incidence of unmatched records much lower than what actually exists in the database. This provides a view of the algorithm effectiveness in discovering matched pairs. Algorithm tuning may be required to reduce the incidence of unmatched pairs.

\[ FNR = \frac{\text{False Negative 'non-match' Pairs (FNMP)}}{\text{Total EMPI Matches (TEM)}} \times 100 \]

3.7 Indeterminate Match Rate (IMR): This is also known as ambiguous match rate. Of the total number of evaluations performed by the algorithm (TEM), the percent that were found to be indeterminate matches after validation. These are matches where the pair of candidate records offered by the algorithm did not have sufficient information to make a clear determination of whether or not they were the same individual. This may indicate a number of different factors such as data quality or data capture challenges, business process variation, “old” data that was not adequately managed, or simply business factors that influenced where an organization sets a matching threshold.

\[ IMR = \frac{\text{Indeterminate Match Pair (IMP)}}{\text{Total EMPI Matches (TEM)}} \times 100 \]

3.8 Matching Accuracy Rate (MAR): This is the overall accuracy rate of the demographic matching process over a given period of time.

\[ MAR = \frac{\text{Total Match Pair (TMP) + Total NonMatch Pair (TNMP)}}{\text{Total EMPI Matches (TEM)}} \times 100 \]
3.9 Matching Error Rate (MER): This is the overall error rate of the demographic matching process over a given period of time.

\[ \text{MER} = \frac{\text{False Positive Match Pair (FPMP)} + \text{False Positive NonMatch Pair (FNMP)}}{\text{Total EMPI Matches (TEM)}} \times 100 \]

4.0 Which Performance Indicators Should I Use?
All or some of these KPI can be used by a variety of audiences. A person’s job responsibilities as they relate to the MPI will drive which KPIs are best utilized for their purposes. The indicators included in this section of the toolkit are important to five major audience categories:

1) C-Suite, upper management, decision and policy makers,
2) Department heads, supervisors, quality assurance monitors, practice managers and other middle management compliance monitors,
3) Front line supervisors and staff, and
4) Health Information Exchanges (HIEs).
5) Vendors and developers of MPI software

Category 1 utilizes the indicators for strategic planning, financial management and high level decision and policy making. Category 2 includes department heads, quality assurance monitors, practice managers and other middle management monitoring activities. Category 3 relates to the front line employee and supervisors to monitor daily performance and perform root cause analysis of day to day operational issues. These are usually found in Access Management, HIM, IT, among other areas.

The following discusses each category of audience and how they may use these performance indicators.

4.1 Category 1: C-Suite, Finance, Upper Management and Decision & Policy Makers
There are two Key Performance Indicators that this audience needs to master. They are: 1) the EMPI Database Duplicates (EDD) and 2) the EMPI Database Duplicate Percent (EDD%). The EDD is the numeric count of pairs of records identified by the EMPI or system algorithms that represent potential duplicates or multiples within the organization’s database.

For instance, an EMPI contains 1,000,000 lines or records. The algorithm identifies 125,000 potential pairs (250,000 records). This results in an EDD equal to 125,000 pairs (potential duplicates) and an EDD% of 12.5%. This is computed by dividing number of identified pairs (125,000) time 100, by the total number of lines in database (1,000,000).

\[
\text{Number of Identified Pairs} \times 100 \quad \text{or} \quad \frac{125,000 \times 100}{1,000,000} = 12.5\% 
\]

The national average for database duplicates ranges from 8 to 12% with larger databases often as high as 30-40%. There are hospitals in the United States who, through proactive duplicate remediation, have reduced their duplicate rate to less than 0.002%.

Studies have shown the hidden costs of duplicate records to be substantial. These two KPIs have major impact on three areas: 1) data integrity, 2) clinical order and resulting, and 3) the revenue cycle. Data should be viewed as an organizational asset. The quality and integrity of the data are critical to the day to day operation. Quality data allows an organization to leverage that asset for strategic planning, product and service development, among others. It becomes a critical success factor. Inaccurate data impacts negatively on decision making authority, compliance with standards, unnecessary work and rework, not to mention patient misadventures. Incomplete or inaccurate information elongates the revenue cycle due to potential denials or payment delays. It creates cash flow problems due to compromised information delaying invoicing or claims processing and increasing days in accounts receivable.

4.2 Category 2: Department Heads, Supervisors, Quality Assurance Monitors, Practice Managers and Other Middle Management Compliance Monitors.
In addition to the EDD and the EDD%, middle management should understand all nine of the performance indicators. Department heads, supervisors, quality assurance monitors, practice managers and other middle management compliance monitors use these performance indicators as benchmarks against which to measure the quality of the data but also the performance of staff, efficiency of the procedures and accuracy of the algorithm. Deviations in these metrics may indicate the need to re-optimize matching parameters and may also trigger the need for an EMPI database cleansing exercise.

4.3 Category 3: Front Line Supervisors and Staff
This category of employee needs to master all nine performance indicators. Success in business relies on making the right decisions at every level. Organizations and executives focus on high-impact, strategic decisions. Operational decision making is often neglected because the impact of individual front-line decisions is frequently incremental and cumulative in their business impact. It is important to empower employees with the relevant information they need to make better, more strategic decisions. PII KPI metrics provide the front line supervisors and staff with relevant information to support the performance management processes that effectively communicate organizational aligned goals. Accomplishing organizational goals relies on the accuracy and quality of the data as well as ongoing employee education, training and performance evaluation.

Daily patient identity integrity activities of front line supervisors and staff require all to have good understanding and regular utilization of the KPI metrics noted above.

4.4 Category 4: Health Information Exchanges (HIEs).
This category needs to master all nine performance indicators. System monitoring is vital to HIEs’ PI Integrity. Although many HIEs are in the early stages of implementations most HIEs have the ability to log matching decisions. These processes are used for the ongoing evaluation and monitoring of matching performance. These KPIs begin to form a framework for performance measurement, which includes technology, human resources and workflow. Anomalies in matching accuracy need to be subjected to root cause analysis to determine appropriate corrective actions.
Also important to ensuring accurate record linking is the need for HIEs to determine the EMPI record linking thresholds and include these thresholds requirements as part of their data governance and data sharing agreements. After the HIE has had a chance to profile data quality issues, develop a tracking, reporting, and resolution of data issues process the PII KPIs become an integral part of the data quality management efforts. EMPI data thresholds set the bar for acceptable quality and integrity of the match of records across the exchange.

4.5 Category 5: Vendors and Developers of MPI Software

Vendors and developers of MPI software provide more value to their customers through the provision of meaningful and effective management and reporting tools. Providing users with self-service analytics and reporting capabilities, facilitates the operational requirements of the client, and improves the quality and efficiency of data and organizational processes for all systems.

PII KPIs contribute to a standardized approach to defining, deploying, and managing MPI data to help ensure reporting consistency across systems and deliver faster, more accurate results across the enterprise.

When requesting proposals from vendors, providers must include in their vendor requirements documents these basic functionalities for their backend system reporting. Vendors should be required to provide these as basic system functions and not as a “custom report” at additional cost.
Appendix A: Glossary Terms

**Algorithm**: A step-by-step procedure utilizing mathematical calculations and predefined rules. Algorithms are used for calculation, data processing, and automated reasoning.

**Community Patient Index (CPI)**: Is a population registry, used for health care purposes. The CPI number uniquely identifies a person on the index.

**Deterministic record linkage**: The simplest kind of record linkage, called deterministic or rules-based record linkage, generates links based on the number of individual matching identifiers. The technique does an exact match of data elements and assigns weights, or significant values to particular data elements and then uses these weights in the comparison of one record to another.

**Duplicate**: Condition where one patient has two or more different medical records within the same organization. This may be caused by the existence of more than one unique identifier (e.g., medical record number or person identifier) for the same person in the MPI.

**EMPI Matching Validation (EMV)**: An analysis of the Total EMPI Matches (TEM) which identifies the following: total number of true match pairs, false positive matched pairs, false negative match pairs and indeterminate match pairs.

**False Match**: The system incorrectly matches a pair of records when they are not a match but instead represent data concerning two separate individuals.

**False Match Rate**: Measures the percent of invalid matched pairs that are incorrectly accepted as matched pairs.

**False Positive**: A match result that indicates that two records represent the same person when in reality they do not. Two individuals’ records are incorrectly declared to be a match.

**False Negative**: A match result that is erroneously negative. Two records for the same person are thought to relate to separate individuals when in reality they both concern the same person.

**Indeterminate Matched Pair (IMP)**: These may also be known as ambiguous matches. These are matches where the pair of candidate records offered by the algorithm do not have sufficient information to make a clear determination of whether or not they are the same individual. This provides an indication of poor data quality in data collection and field completion or business decisions related to threshold setting for algorithms.

**Key Performance Indicator**: Predetermined quantifiable measurements that reflect the critical success factors of an organization. KPIs enable management to track the matching performance of their EMPI processes.

**Linking Records**: Linking is a means of associating records in an EMPI dataset to indicate the records refer to the same person across different data sources.
Master Data Management (MDM): The business processes and technology to support the mastering of core data that supports critical business processes across an enterprise. The data are at the heart of business transactions. In healthcare, this would be key data about the patient, provider, location, insurer, households, reference codes, and more.

Master Person (Patient) Index (MPI): A master person index (MPI) is an electronic demographic database that holds information on every patient registered at a health care organization.

Enterprise Master Patient/Person Index (EMPI): An Enterprise Master Patient/Person Index is a database that contains a unique identifier for every patient/person in the enterprise as well as tables connecting the EMPI identifier to the identifiers in all registration systems. All registration systems would look to the EMPI to obtain patient information based upon several identifiers. Healthcare organizations or groups of entities implement enterprise-wide EMPIs to identify, match, merge, de-duplicate, and cleanse patient records to create a master index that may be used to obtain a complete and single view of a patient.

Match: Two or more records in a database identified as a result of a comparison process or algorithm where all or some identifiers meet predefined rules to suggest a probability of being the same individual. A match at this level is not necessarily a true match. Some may require validation to obtain certitude. Common identifiers used in matching comparison are name (full, middle, last), date of birth, and gender. Relying solely on names, date of birth and gender is insufficient to validate that the matched records belong to the same individual.

Matching Algorithm: An algorithm used by an EMPI to determine the probability that two separate data sets represent the same or different individuals.

Multiple: More than two records are offered as match candidates for one individual. Multiples are processed two at a time as pairs until all are resolved.

Non-Match: A pair of records that are determined NOT to be the same individual. The determination may be based on a predetermined probabilistic threshold based on the organizations agreed upon matching criteria.

Overlap: One individual having more than one record within and/or across facilities in an enterprise (e.g., patient Sam Jones has medical record number 54321 at Facility A and medical record number 4887733 at Facility B).

Overlay: Two or more distinct individuals erroneously assigned the same identifier (e.g. medical record number) and thus co-mingling their health information into one record.

Pair: Two records offered as a potential match.

Probabilistic Record Linkage: A statistical linking process, sometimes called fuzzy matching, that takes into account a wide range of potential identifying data, computes weights for each identifier based on its estimated ability to correctly identify a match or a non-match, and uses these weights to calculate the probability that two given records refer to the same individual.
Threshold: The organization’s predetermined level of acceptance of an algorithm’s match validity. The thresholds define the range of scores acceptable to the institution for an exact match, high-medium-low match, and no match.

Positive match threshold (PMT) - The EMPI evaluation positive match threshold.
Negative match threshold (NMT) - The EMPI evaluation non-match threshold.

Validate: A process used to authenticate, verify, or prove that a candidate pair is in fact the same individual. Validation indicates a level of review that provides official sanction.