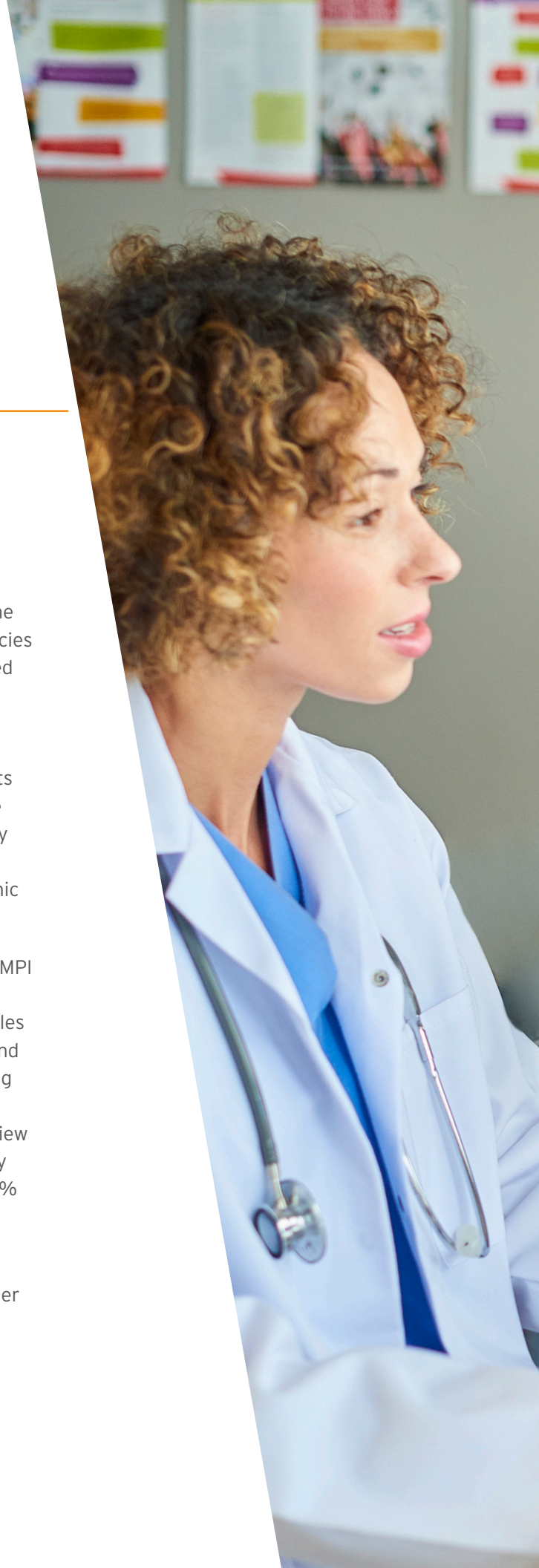
ARE YOUR ANALYTICS PUTTING YOUR MPI DATA INTEGRITY - AND QUALITY OF CARE - AT RISK?

IRON MOUNTAIN®

ANALYZING THE INTEGRITY OF MASTER PATIENT IDENTITY MATCHING ALGORITHMS



INTRODUCTION

As the healthcare industry advances information governance best practices and accelerates value-based care transformation, the strategic importance of data quality and integrity is becoming increasingly apparent. In fact, this topic recently garnered public attention with the filing of a lawsuit in excess of $1 billion dollars[1]. The lawsuit alleges potential harm of millions of patients due to deficiencies in the way patient information was managed, identified and displayed in the EMR system. Key deficiencies cited include the display of incorrect medical information as well as the erroneous display of multiple patient's information, which are alleged to have resulted in numerous patient care errors and mass breaches of privacy. Lawsuits such as this one point to the immense and highly complex challenge providers – and EMR vendors – face as they seek to address accuracy of electronic patient information in the most critical databases and fundamental systems, such as the Master Patient Index and Electronic Medical Record.

Clearly, it is not for a lack of trying. Many providers seek to address MPI integrity issues devising data entry and registration quality controls and deploying processes that leverage EMR-embedded report modules or standalone MPI analytics software to identify errors, duplicates and overlays on a routine basis. Sadly, while many spend hours analyzing and triaging the findings, few have done the diligence to ensure the algorithms behind the analytics software offer the comprehensive view and strategic weighting required to deliver the high level of accuracy needed to enhance integrity and achieve best-in-class error rates (2% or less).

Understanding that many of the details behind the analytics can be mind boggling or just completely overwhelming, we've pulled together a few easy-to-follow examples of what to look for when evaluating matching alogrithms as well as additional details around why it matters.

1 https://www.healthcareinfosecurity.com/1-billion-lawsuit-focuses-on-ehr-data-integrity-concerns-a-10463?rf=2017-11-21_ENEWS_SUB_HIS_Slot1&mkt_tok=eyJpIjoiTURRNE4yUmxZemsxWW1NeCIsInQiOiJicHIJdVFqeWZZSWh6SG1UeURobGY5am9Dc3dydU9Oa25tTEdzQVVFV3A2QXIIZDROOGltOFVpYTFKS3FjakZmZVZRU1ZJdVFZeHJLMSt3M1hcL3V6REtKZnRYXC9OSWFTZ1VoYWtraVNJNUhKeUtxTnBSRUFrcUFMZDBQSjY2bFpWIn0%3D

CASE STUDY

HOW DO WE KNOW THE ALGORITHMS BEHIND THE ANALYTICS MATTER?

Leveraging a frequently used reporting module, one healthcare provider identified 1 million duplicate pairs (or errors) in their EMR system. The EMR report module deemed 625,000 pairs eligible for automated matching based on the weight for each match being above a specified threshold. In other words, the system was saying that each matched pair was the same person and that it was appropriate to merge the patient records without additional analysis or human intervention. However, the same data analyzed by a purpose-built software using a probabilistic algorithm yielded very different results. Of the original 625,000 pairs deemed acceptable for automatic merge, the probabilistic software determined:

- 18,279 did not meet the automated matching threshold and required human review
- 357 were not worth reviewing (meaning it was very clear that the records were not for the same person)

Just imagine the risk your institution would be exposed to had you merged the records based upon the guidance from the EMR report module and a patient had an adverse outcome after being treated by your clinicians.

The probabilistic algorithm also uncovered an additional 295,000 pairs that actually were eligible for automated matching that the original EMR report module failed to identify. Since they were not auto-matched when they could have been, the institution incurred additional costs to manually review and determine the appropriate disposition for these records when it was not necessary.

Unfortunately, this scenario is not uncommon. While not widely known and seldom discussed, the reality is the integrity of your MPI data is only as good as the algorithms behind the software running the analysis. The good news is there is a simple way to determine if you have selected the right analytics software.

THREE QUESTIONS THAT CAN HELP EVALUATE YOUR MPI ANALYTICS REPORTING TOOLS

1 IS MY SOFTWARE OR REPORTING MODULE RUNNING A PROBABILISTIC OR RULES-BASED ALGORITHM?

Algorithms are the foundational component for any analytics software. Using the wrong rules, an incomplete set of requirements or rules existing outside a probabilistic lens, significantly and severely hinders your ability to accurately identify duplicates and errors in your system.

STANDARD PRACTICE: RULES-BASED ALGORITHM:

A rules-based algorithm requires "tuning" by the organization in order for the algorithm to work. What this means is that you assign a weight to each element in the matching logic (e.g. Name, Date of Birth, etc.) and then multiple testing runs are completed to review the weights. The weights may also be adjusted multiple times before the algorithm is put into production and, even then, it is only as accurate as the human-derived weighting. So bottom line, rather than relying on industry experts with years of experience in proper weighting of elements and a sophisticated algorithm that's been tested and improved over time, you must be the expert.

BEST PRACTICE:

Your analytic software should use a probabilistic algorithm which does not require tuning by you and your organization. A probabilistic algorithm automatically determines how much weight each data element should receive in a match by statistically determining the quality of that element and how often it is present within the data. Default or bogus data – such as a social security number (SSN) of 999-99-9999 - is flagged within the algorithm so that it is not treated as a valid match.

WHY IS THIS IMPORTANT?:

It is not uncommon for false positives to exceed the automatic match threshold with a rules-based algorithm, meaning that pairs who do not represent the same person are automatically merged or linked together. One reason is that people tend to believe that a social security number is unique and they will often assign a high weight to it in a rules-based algorithm. If the assigned weight of the SSN is high enough, then the positive weight of a matching SSN will overcome the negative weight associated with a mismatch on other demographic data. This concept is depicted in Table 1 below. In Table 1, you see an example of a pair that will be automatically merged, even though it is a case of a husband and wife who both have the same SSN listed in the registration system. While the SSN really is unique in the sense that it is assigned to only one person, that doesn't mean that it is entered into the registration system for the correct patient every time.

TABLE 1

MRN	LAST NAME	FIRST NAME	D.O.B.	SSN	ADDRESS
123	SCHULTZ	RAYMOND	05-01-1960	432-67-8768	123 MAIN STREET
687	SCHULTZ	MELISSA	08-12-1960	432-67-8768	123 MAIN STRET

2 HOW ARE VALUES WEIGHTED?

How values are weighted is important. This determines what fields identify duplicates and what scale will be used to determine duplicates eligible for auto-merge and those that require manual review.

STANDARD PRACTICE:

Many reporting tools will weigh fields the same, without taking into consideration the uniqueness of the field, the frequency of use or whether the element is subject to change over time. These same tools set a high threshold for manual review, meaning a large population of records will be recommended for automatic merging that could, and should, be reviewed manually. This results in high instances of matching the wrong records.

BEST PRACTICE:

Your analytics tool should be sophisticated enough to weigh matches with consideration of specific demographic criteria. For example, unique values should be granted a higher weighting and demographic elements that are consistent (e.g., date of birth) should be weighted higher than those that change (e.g., address). Additionally, the threshold for automatic merging should be relatively strict, isolating only those records that are strategically weighted, with a high probability of being duplicates and where certain conditions are met.

WHY IS THIS IMPORTANT:

Most vendors perform automated matching based on weight threshold alone and do not specifically look for conditions that warrant human review. The example of the husband and wife match in the first question highlights how a pair might represent two different people even though the weight of the pair is above the threshold for automated matching. A combination of a weight threshold and specific matching rules provides a better outcome with less risk to the organization. In the husband and wife example, it should be flagged for manual review because the genders and dates of birth do not match. A variation on even one of these elements significantly increases the risk of a matching error.

YOUR ANALYTICS TOOL SHOULD BE SOPHISTICATED ENOUGH TO WEIGH MATCHES WITH CONSIDERATION OF SPECIFIC DEMOGRAPHIC CRITERIA.

3 HOW ARE DUPLICATES COUNTED?

How duplicates are counted is critical to effective management and efficient resourcing of your MPI cleanup work streams. What's more, it's critical to calculating your error rate accurately.

STANDARD PRACTICE:

Many EMR reporting tools will count duplicate records multiple times. This concept is depicted in Table 2. The letters in Table 2 represent unique records in the system. The numbers across the bottom represent the total number of duplicates identified by the software based on the logic behind the analysis. In this case, the EMPI Vendor analytics identified 5 unique records, 4 of which are duplicates. However, the EMR Software analytics identified 10 total duplicates, counting each possible pair that can be made from the 5 unique records as a possible pair.

TABLE 2

COUNTING MULTIPLES					
EXAMPLE: PATIENT WITH 5 UNIQUE RECORDS					
EMPI VENDOR ANALYTICS		EMR SOFTWARE			
A					
B		AB			
C		AC	BC		
D		AD	BD	CD	
E		AE	BE	CE	DE
5		4	3	2	1
TOTALS					
4		10			

WHY 4 TOTAL INSTEAD OF 5?

The total is four because of the five records four are duplicates and one is the surviving record.

Taking this a step further, if a customer is bringing together records from two different EMR systems, the pairs that involve one record from system 1 and one record from system 2 will get counted twice; once as a system 1 duplicate and once as a system 2 duplicate, which further inflates the duplicate volume and rate.

BEST PRACTICE:

Your analytics software should be able to identify and isolate the unique records associated with a patient to give you a clear view of the work effort required to merge each instance of a single patient's records and to avoid having you and/or your team do unnecessary manual review cycles.

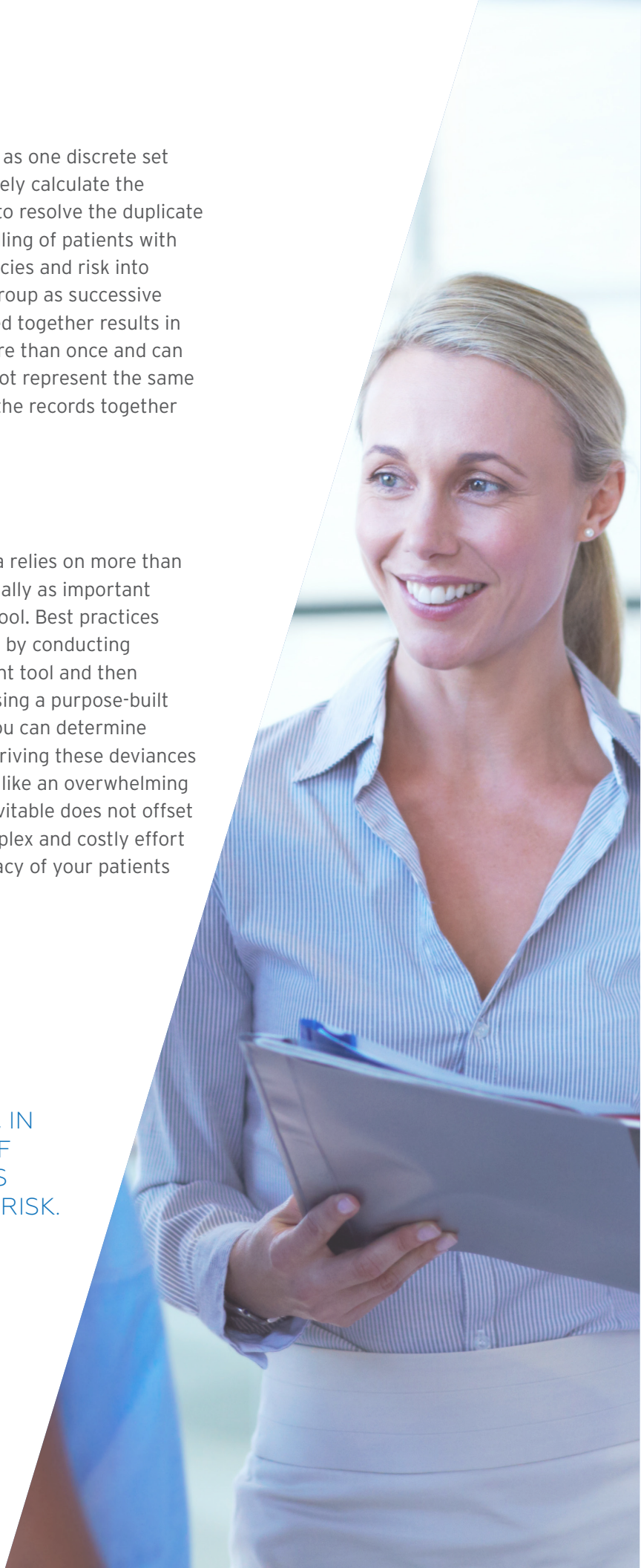
WHY IS THIS IMPORTANT?:

Systems that cannot show the multiple group of records as one discrete set double count records, making it more difficult to accurately calculate the duplicate rate and determine the work effort necessary to resolve the duplicate records in the group. Additionally, poor display and handling of patients with more than two potential records can introduce inefficiencies and risk into the duplicate resolution process. Showing the multiple group as successive pairs of two rather than one group that must be managed together results in unnecessary work, as the same record is researched more than once and can result in errors in confirming that the records do or do not represent the same person. This is because the user is not able to see all of the records together and assess their relationship to each other cleanly.

CONCLUSION

As the examples illustrate, the accuracy of your MPI data relies on more than simply having analytics software or a reporting tool. Equally as important is to vet and understand the algorithms underlying the tool. Best practices recommend analyzing the effectiveness of analytic tools by conducting a comparative analysis of MPI integrity using your current tool and then conducting the same analysis on the same set of data using a purpose-built analytics tool. In comparing the findings to your own, you can determine where the discrepancies reside, explore the root cause driving these deviances and build a plan to address the gaps. While this can feel like an overwhelming task, it is imperative to be proactive. Prolonging the inevitable does not offset the need. Rather, it will result in an inherently more complex and costly effort while, in the interim, placing the quality of care and privacy of your patients increasingly and significantly at risk.

> PROLONGING THE INEVITABLE DOES NOT OFFSET THE NEED. RATHER, IT WILL RESULT IN AN INHERENTLY MORE COMPLEX AND COSTLY EFFORT WHILE, IN THE INTERIM, PLACING THE QUALITY OF CARE AND PRIVACY OF YOUR PATIENTS INCREASINGLY AND SIGNIFICANTLY AT RISK.



800.899.IRON | IRONMOUNTAIN.COM

ABOUT IRON MOUNTAIN

Iron Mountain Incorporated® (NYSE: IRM) is the global leader in storage and information management services. Trusted by more than 220,000 organizations around the world, Iron Mountain's real estate network comprises more than 85 million square feet across more than 1,400 facilities in 46 countries dedicated to protecting and preserving what matters most for its customers. Iron Mountain's solutions portfolio includes records management, data management, document management, data centers, art storage and logistics, and secure shredding, helping organizations to lower storage costs, comply with regulations, recover from disaster, and better use their information. Founded in 1951, Iron Mountain stores and protects billions of information assets, including critical business documents, electronic information, medical data and cultural and historical artifacts. Visit www.ironmountain.com for more information.

© 2018 Iron Mountain Incorporated. All rights reserved. Iron Mountain and the design of the mountain are registered trademarks of Iron Mountain Incorporated in the U.S. and other countries. All other trademarks and registered trademarks are the property of their respective owners.

USHC-BP-050718A