ONC NPRM (Feb 2019): Patient Matching RFI (p. 7555)

Public health has significant experience over a long period of time in patient matching strategies for records collected from diverse clinical locations. The following observations and suggestions are offered based on this experience:

RFI Question	Response
It is a common misconception that technology alone can solve the problem of poor data quality, but even the most advanced, innovative technical approaches are unable to overcome data quality issues. Thus, we seek input on the potential effect that data collection standards may have on the quality of health data that is captured and stored and the impact that such standards may have on accurate patient matching. We also seek input on other solutions that may increase the likelihood of accurate data capture, including the implementation of technology that supports the verification and authentication of certain demographic data elements such as mailing address, as well as other efforts that support ongoing data quality improvement efforts.	The quality of data used for patient matching is indeed a difficult problem which has plagued public health registries for some time. As we described in an article published in 2017, ONC convened a Patient Matching Community of Practice in 2014-15. We wrote, "Its major focus was developing a five-level data quality maturity model to try to characterize an organization's sophistication in using different common data elements to perform patient matching functions, as well as articulating value propositions for improved matching for different stakeholder types. The project released two documents, _Developing and Testing a Data Management Model and Maturity Scale Tailored to Improving Patient Matching Accuracy_ and Guidelines for Pilot Testing of Data Management Maturity℠ Model for Individual Data Matching describing its work. The Data Quality Maturity Scale, included as Appendix B, highlights how systems across the healthcare community, at least as reflected in the core data elements, are at the high levels of maturity. In practice, however, the data elements needed for levels 4 and 5 are precisely the ones that are least consistently captured." We encourage ONC to draw on these documents and resources whose development ONC funded. External validation of key data elements used for matching can also be a big help. For example, in 2017 the American Immunization Registry Association (AIRA) arranged access to SmartyStreets, a cloud-based address cleansing service, for all Immunization

RFI Question	Response
	Information Systems (IIS) which chose to access it. By leveraging available CDC funding, for a modest amount this service is able to cover the *entire* IIS community and significantly increase the level of quality in address data which is often key for proper patient matching. AIRA maintains the license, provides documentation and coordination, and sponsors a monthly user group of interested IIS projects.
In concert with the GAO study referenced above, we seek input on what additional data elements could be defined to assist in patient matching as well as input on a required minimum set of elements that need to be collected and exchanged. We encourage stakeholders to review the Patient Demographic Record Matching section of the Interoperability Standards Advisory and comment on the standards and implementation specifications outlined. Public comments and subject matter feedback on all sections of the Interoperability Standards Advisory are accepted year round.	The Patient Demographic Record Matching Sections seems inadequate to address data elements for patient matching as it primarily focuses on IHE transactions which do not seem to focus normatively on *which* data elements might be best for matching. The Data Quality Maturity Scale, included as Appendix B in Guidelines for Pilot Testing of Data Management Maturity℠ Model for Individual Data Matching referenced above, provides detailed suggestions for data elements to be used for patient matching that were vetted through the community of practice that developed the guidelines. In addition, in January 2019 AIRA published its *IIS Functional Guide, Vol. 2: CDC Endorsed Data Elements*. This exhaustive document includes (in Appendix C) a list of data elements endorsed to fulfill the IIS functional standard of identifying, preventing and resolving duplicated and fragmented patient records using an automated process. This list is also worth consulting. Research in New York City by the Citywide Immunization Registry (CIR) has demonstrated that though matching is a complex activity, and it is difficult to tease apart factors affecting successful matching, the search success rate for the CIR was higher when more

RFI Question	Response
	search fields were sent, especially the internal ID assigned to each patient in the CIR and available to EHRs that query the system should they choose to store it. Studies such as this one should be replicated to help determine the most effective fields for searching and matching.
Also in alignment with the GAO study, we seek input on whether and what requirements for electronic health records could be established to assure data used for patient matching is collected accurately and completely for every patient. For instance, the adopted 2015 Edition ''transitions of care'' certification criterion (§ 170.315(b)(1)) currently includes patient matching requirements for first name, last name, previous name, middle name, suffix, date of birth, address, phone number, and sex. These requirement also include format constraints for some of the data.	Requiring specific data quality for is admirable but may not be practical, since in many (if not most) cases an EHR can only contain data as good as what is provided by the patient. To the degree that data formats can be enforced (like data formats for date of birth), or standard value sets maintained (like sex, race, and ethnicity), the quality of the data will naturally improve.
There are unique matching issues related to pediatrics and we seek comment on innovative and effective technical or non-technical approaches that could support accurate pediatric record matching.	The IIS community has worked in this domain specifically for more than twenty years. There are a number of specific patient matching issues that affect pediatric records, including: • Birth records that do not contain a true first name (but rather are populated with "baby boy" or "baby girl" as a first name was not available) can become difficult to match to future records. • Multiple births can sometimes present confusing matching problems, especially when first names are close or even identical. • Children do not usually have records in referential matching databases that are primarily drawn from financial/credit data sources (see below). • Though not unique to children, some data sources may include a patient's middle name embedded in the patient's first name field.

RFI Question	Response
	• Children may lack common identifiers that adults typically possess that may be used as primary or secondary matching fields (*e.g.,* driver's license number, social security number, cell phone number, e-mail address, unique Medicaid ID [which may be a family ID]). • On the other hand, children are often associated with parents/guardians and parent/guardian data can be used to supplement primary data for matching. There are no magic answers to addressing these issues; technology developers need to be sensitive to them when crafting solutions to pediatric matching challenges.
Recent research suggests that involving patients in patient matching may be a viable and effective solution to increase the accuracy of matching, and giving patients access to their own clinical information empowers engagements and improved health outcomes. We seek comment on potential solutions that include patients through a variety of methods and technical platforms in the capture, update and maintenance of their own demographic and health data, including privacy criteria and the role of providers as educators and advocates.	Public health registries are only just beginning to provide direct access to patients; IIS are probably leading the way given the broad usefulness of an up-to-date immunization history and forecast for school/child care/camp admission and preventive care. Many IIS also perform automated or semi-automated outreach services to encourage patients to complete missing immunizations ("Reminder") or to warn them of upcoming immunization needs ("Recall"). These services will often use text messages or e-mails to contact patients directly, yet IIS often do not have complete cell phone or e-mail records for their patients. Most IIS projects are somewhat reluctant to accept patient contact information (which could then also be available for matching) directly as opposed to soliciting this information from healthcare providers when they submit immunization records *to* the IIS. We do feel there is some potential for augmenting IIS contact information with patient-supplied data once patient access to IIS data becomes more prevalent.

RFI Question	Response
In addition, we seek input on standardized metrics for the performance evaluation of available patient matching algorithms. Health IT developers are each relying on a number of patient matching algorithms, however, without the adoption of agreed upon metrics for the evaluation of algorithm performance across the industry, existing matching approaches cannot be accurately evaluated or compared across systems or over time.	This has always been a difficult topic and we do not see any easy answers ahead. In 2017 ONC sponsored the Patient Matching Algorithm Challenge (PMAC) whose was to allow vendors to compete for the highest performance metrics for their matching algorithms by testing their software against a large set of test data provided by ONC. Cash prizes were awarded in a number of categories, and the winning vendors were featured in the discussion on the webinar. One of the main purposes of the challenge was to promote the use of standard metrics to evaluate algorithm products. We were a little concerned that the winners by their own admission "analyzed patterns in the data." This seems to call into question the applicability of their results to the "real world" where you don't get to see the data set; you have to adjudicate them as they come in. That means that these particular test runs were "tuned" for the data set and the measurable results might not hold up for other data sets. Over the years, several public health initiatives have attempted to provide comparative measures of matching algorithm performance or quality and have had less than successful results.
At the same time, we seek input on transparent patient matching indicators such as database duplicate rate, duplicate creation rate, and true match rate, for example, that are necessary for assessment and reporting. The current lack of consensus, adoption, and transparency of such indicators makes communication, reporting, and cross- provider or cross-organizational comparisons impossible, impedes a full and accurate assessment of the extent of the problem, prohibits informed decision making, limits	We have no comment on this important question.

RFI Question	Response
research on complementary matching methods, and inhibits progress and innovation in this area.	
There are a number of emerging private-sector led approaches in patient matching that may prove to be effective, and we seek input on these approaches, in general. A number of matching services that leverage referential matching technology have emerged in the market recently, yet evaluations of this type of approach has either not been conducted or has not been made public. Other innovative technical approaches such as biometrics, machine learning and artificial intelligence, or locally developed unique identifier efforts, when used in combination with non-technical approaches such as patient engagement, supportive policies, data governance, and ongoing data quality improvement efforts may enhance capacity for matching.	In an article published in 2017, we identified a set of distinct strategies for matching that seemed to be in play and the lack of any real consensus around any of them: 1. A "traditional" approach which leverages either deterministic and/or probabilistic techniques that continue to struggle with the lack of standardized data for input as discussed elsewhere in this response; 2. A unique identifier approach, either government sponsored or managed by the private sector, though this would likely be insufficient without corroborating data in a population as large and diverse as the US; 3. Health record banks which put the patient at the center of the problem but which have failed to gain any traction in the marketplace; 4. Biometrics, which still suffer from some limitations as well as privacy concerns; 5. Newer, innovative approaches such as referential matching which still have limitations in some segments of the population (like children) We believe that the public and private sectors need to get together to discuss and pilot various approaches and to encourage Congress to reexamine its current position on a national unique patient identifier.
Finally, ONC seeks input on new data that could be added to the United States Core Data for Interoperability (USCDI) or further constrained within it in order to support patient matching.	Refer to the Data Quality Maturity Scale, included as Appendix B in Guidelines for Pilot Testing of Data Management Maturity℠ Model for Individual Data Matching which provides detailed suggestions for data elements to be used for patient matching that were vetted through the community of practice that developed the guidelines.