CLIAC Public Comment April 10, 2024

Thank you for the opportunity to offer public comment as a citizen and laboratory informaticist. I am clinically trained as a medical laboratory scientist with experience in academic medical centers to smaller clinic settings especially in different needs therein. My PhD is in Health Informatics and I have a passion for laboratory data interoperability and usability of laboratory data for a variety of clinical, public health and research purposes. I'm also the first laboratory professional who is a Fellow of the American Medical Informatics Association.

Regarding the topic of Artificial Intelligence (AI) and Machine Learning (ML), I want to thank Dr. Carter for her presentation on these topics that are gaining in popularity. My doctoral training in Health Informatics includes courses in Artificial Intelligence which includes AI and ML methods in Dr. Carter's presentation, Clinical Decision Support including Tools and Impacts on Decision Making, Healthcare Data Standards to name a few. These tools can be utilized for many great applications, as well as cause harm or bad decisions if not designed/set up correctly or "hallucinate" with black box outputs. I created a simple neural network to classify anemia based upon common complete blood count parameters for a course project so these are easy to develop by many, including those without laboratory expertise.

As applications flood the market or even in those applications that one may create or customize, the question becomes how do we know these tools are functioning as expected across different care settings, sources of data, patient populations, etc. and safely, not resulting in bad outputs, patient harm, or bad decisions or recommendations? They need to be clinically validated that they are "fit for purpose" and generate quality data and outputs. I want to emphasize one of Dr. Carter's points that **quality of data is critical**. It's critical to have good data to avoid GIGO: garbage in, garbage out.

As CLIAC deliberates what may be needed to support their safe use, one consideration is where and how are they being used? Are they low risk decisions like a spell checker in software or high risk decisions such as patient care aspects? Are we even aware where they are being used within software products, whether health IT (EHR, LIS) or ancillary systems, software on IVD devices, etc.?

In my work in standards development, specifically with HL7 FHIR standards, discussion has occurred on how traceability of laboratory data and decisions occur in laboratory workflows. For example, autovalidation tables are often set up within a LIS or middleware

to autoverify and release results that are “normal.” The LIS distinguishes between human verified results and those that are autovalidated by the “software.” This provides traceability for root cause analysis and other quality needs and governance.

With Point of Care testing, how do we indicate that a human, either a consumer or health professional performed a test, versus those with companion applicaition usually on a smartphone that “interprets” result values? It may be in conjunction with a camera reader for a urine dipstick result value or a calculation within the device. These may not “visible” to the consumer or health professional buried within the software application or smartphone device itself.

Are these test results comparable if no human intervention is used in their interpretation? Trust of consumer performed and perhaps some health professional performed testing is a concern across the health ecosystem. Are we more apt to trust those without human intervention that may be used in health professional decision making? For consumer performed testing, preanalytical aspects such as specimen quality, as well as performance of the test are all factors impacting trust as downstream users of the data do not know if data quality is compromised if the test was performed on the patient’s pet or another person or invalidated if expired, etc. Consumers may not exercise the same rigor as trained health professionals regarding specimen rejection or acceptability criteria. We know these preanalytical issues can impact test performance and results interpretation whether by humans or machine learning algorithms.

The FDA sponsored Synensys report of the laboratory ecosystem assessed from a systems approach makes the recommendation that laboratory professionals and expertise be involved in many of these informatics based processes so that laboratory needs are considered.

I ask CLIAC to consider what is needed to ensure laboratory data quality and decisions with use of machine learning and make any recommendations to federal agencies to help ensure they are considered by those with AI/ML evaluation processes and regulations.

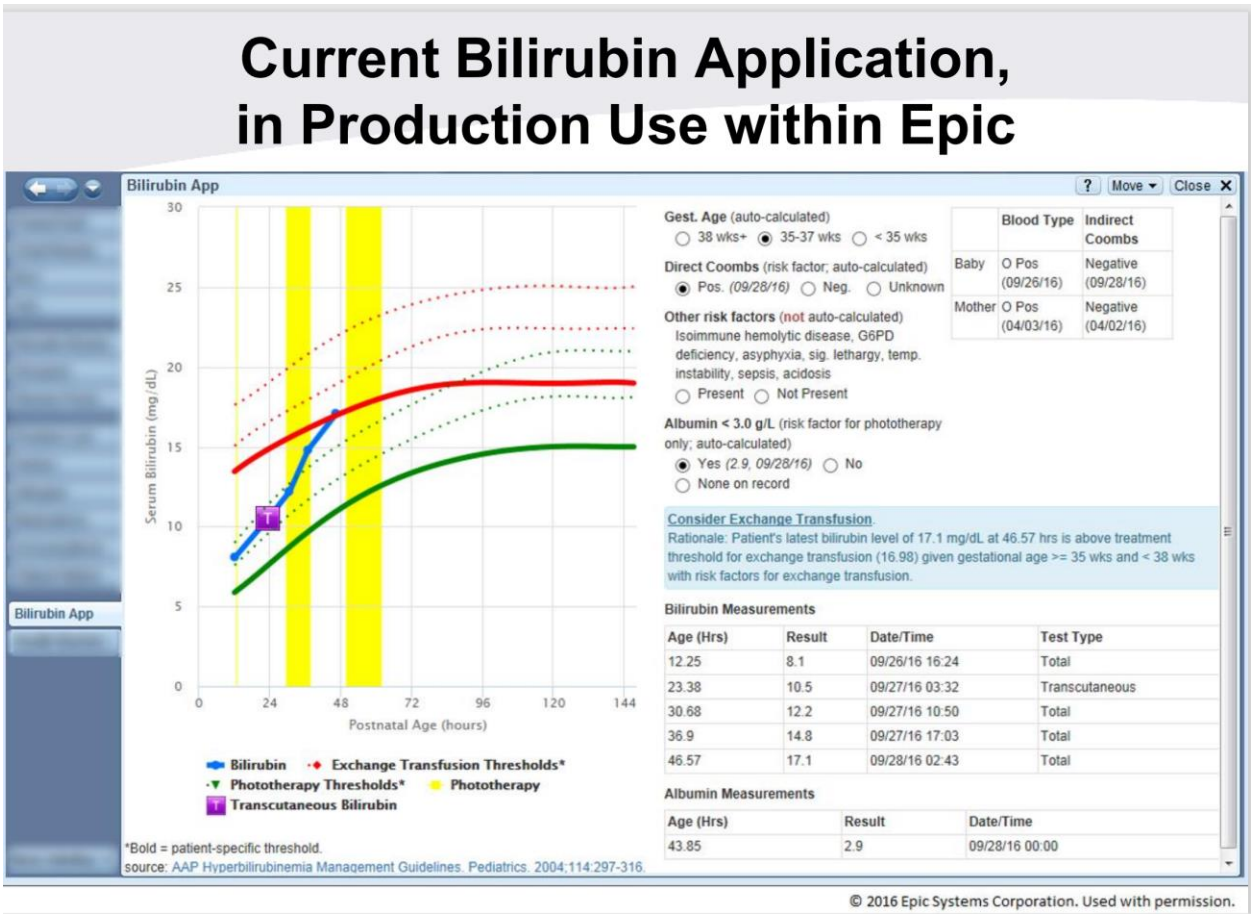
Secondly, with regard to the standards topic. I support standardization and harmonization of laboratory assays and testing to global standards for comparability and interoperability. Those test methods which are not comparable also need to distinguished so all users are aware and do not inadvertently comingle them in a variety of data uses, especially AI/ML. AI/ML trained on lab result values that have clinically significantly different methods/specimens, etc. can introduce bias into algorithms, AI/ML, as well as human assessments and use, as we see happening today. AI/ML will likely magnigy these issues

similar to the transformations we saw when paper based design or data issues were magnified with electronic implementations if they were not addressed.

Harmonization and standardization is a term used by many in the informatics and Healthcare IT space as well. Generally, it is used to mean how do I group the many ways a single lab test is performed into single term or code that I can use to refer to this item, no matter the variety of test names used by each performing laboratory. It is akin to the generic vernacular we use in talking about laboratory tests. We don't usually mention details like specimen or method when we speak about laboratory tests.

Clarity is needed by CLIAC in addressing the Standards questions and options as to which standards might be utilized. Is the focus on standards related to test performance quality as indicated by the FDA standards list showing CLSI documents referenced? Or will standards include terminology/codesystem standards for laboratory data (providing computer processable meaning) and/or data exchange standards used for laboratory data? If the latter, guidance is recommended for how quality implementations and use of the standards will be determined, which use cases or areas of laboratory medicine they will be utilized, and guidance for those inspecting to know especially with newer technologies which are compliant or not? This is an area where caution is needed as the nuances in laboratory testing may not yet warrant a single broad application, but perhaps a smaller scope that may be piloted or phased in. The initial focus may be on simpler areas of testing with lower risk too. CLIAC may also want to recommend that coordination with federal agencies, entities, states and accreditation may be warranted to help ensure definitions, uses, etc of laboratory data are the same/aligned by different entities too.

While I'm not aware of any LIS that has FHIR functionality, much less CLIA Compliant FHIR implementations, there are laboratories using FHIR for ancillary purposes and a vendor who is working on a FHIR LIS. Downstream from the LIS, EHRs and Health IT are certified to meet ONC requirements which includes FHIR and for laboratory data. There is great variability in the quality of laboratory data in these applications. For example, here's a baby bilirubin application where transcutaneous values are listed/graphed with laboratory performed values. (Slide 1 (hl7.org)) On the left scale serum bilirubin is listed, even though a transcutaneous value is not performed on serum. There are many other laboratory data quality issues about this example.



	Blood Type	Indirect Coombs
Baby	O Pos (09/26/16)	Negative (09/28/16)
Mother	O Pos (04/03/16)	Negative (04/02/16)

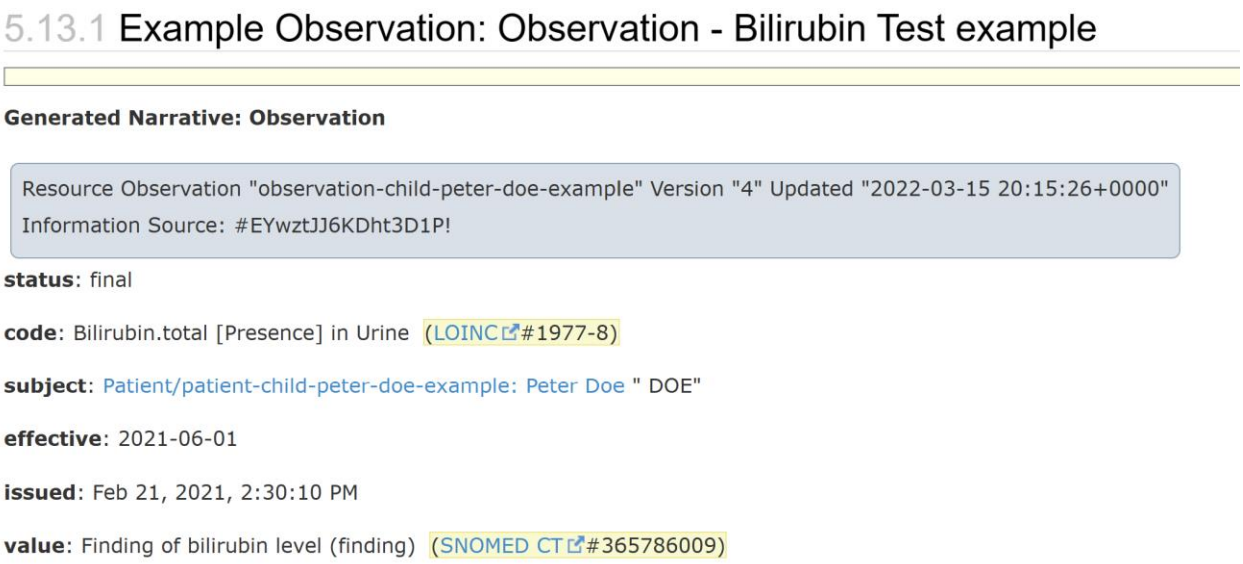
Age (Hrs)	Result	Date/Time	Test Type
12.25	8.1	09/26/16 16:24	Total
23.38	10.5	09/27/16 03:32	Transcutaneous
30.68	12.2	09/27/16 10:50	Total
36.9	14.8	09/27/16 17:03	Total
46.57	17.1	09/28/16 02:43	Total

Age (Hrs)	Result	Date/Time
43.85	2.9	09/28/16 00:00

Consider another example in the screenshot below and link provided, which is named "Bilirubin Test," but is reflected as a qualitative Urine Bilirubin result represented by the LOINC Long Name under "code." (see HL7.FHIR.US.MIHR\Observation - Bilirubin Test example - FHIR v4.0.1) This example doesn't even reflect the lab test name, as it only represents the test with the LOINC (not advised by the LOINC User Guide.) The value is encoded to the wrong SNOMED CT code (it should be from the qualitative value hierarchy). Thus a computer using this code may attribute the wrong meaning to the result value. This test is rarely performed in the US, and there are other LOINCs for the Ictotest and dipstick/test strip methods that provide more detail and clarity. It's unclear whether the effective and issued dates correspond to the specimen collection date, or when the laboratory received the specimen or verified the results in accord with CLIA. The danger is many developers may not know these are important data quality and coding issues and implement these examples "as is," and perpetuate these issues.

A recommendation is for HL7 implementation guides to be developed for laboratory data/use cases with clarity for implementers on how to avoid these issues and have quality implementations. FHIR implementation guides for orders and results be developed with review by CLIA to ensure the end product is compliant with CLIA regulations similar to how

the ONC S&I framework Implementation Guides were developed for lab ordering, (LOI), resulting (LRI), compendiums (eDOS) and ELR Public Health Reporting exchanges. Laboratory expertise is needed in the development of these guides, including from a variety of lab settings to reflect these lab needs too.



5.13.1 Example Observation: Observation - Bilirubin Test example

Generated Narrative: Observation

Resource Observation "observation-child-peter-doe-example" Version "4" Updated "2022-03-15 20:15:26+0000"
Information Source: #EYwztJJ6KDht3D1P!

status: final

code: Bilirubin.total [Presence] in Urine (LOINC#1977-8)

subject: Patient/patient-child-peter-doe-example: Peter Doe " DOE"

effective: 2021-06-01

issued: Feb 21, 2021, 2:30:10 PM

value: Finding of bilirubin level (finding) (SNOMED CT#365786009)

Thank you for your consideration of these comments and helping to ensure the quality and safe use of laboratory and pathology results.

Andrea Pitkus, PhD, MLS(ASCP)CM, FAMIA