Questions for Panelists – Responses of Richard Platt

Learning Health System (LHS)1

1. ***What would you say are the most pressing privacy, data security, and ethical issues regarding the learning health system?***

   Reply:

   We need to appreciate that the best answers are not known for many decisions that patients and providers make every day. This is true despite the great progress we have made. The number of important questions is simply too large to answer using the approaches that have relied on. One important way to understand more about what treatments works best for whom is to learn from the care that is provided every day. The term "Learning Health System" connotes a commitment to improve care, both by learning from all patients' experiences and by implementing the results of the learning activities. On the knowledge development side, the distinction blurs between research, quality improvement programs, and assessment of outcomes involving treatments and medical practices in common use. It is important to distinguish this focus on learning about treatments used in regular practice from research on unapproved, experimental treatments.

   Health records, billing data, and related information are irreplaceable – not simply more convenient -- for answering many critical questions about the effectiveness, safety, and quality of medical care, and to support critical public health functions, such as monitoring the occurrence and spread of serious infections in communities.

   These records are also a critical resource for identifying individuals who might be interested in participating in clinical trials to determine which of several treatments works best for them.

   For these activities to succeed, health records must be accessible for approved uses. Several conditions should apply to the use of this information:

---

[1] *See A Ten Year Vision to Achieve Interoperable Health IT Infrastructure, pp. 2, 8, available at: http://healthit.gov/sites/default/files/ONC10yearInteroperabilityConceptPaper.pdf. See Also, The Learning Health Care System in America, Institute of Medicine of the National Academies, available at: http://www.iom.edu/Activities/Quality/LearningHealthCare.aspx.*

* *Evaluations should use the least amount of identifiable information required to answer a question. Often, all that is needed is a count of the number of people treated and the number with the particular outcome of interest.*

* *Information should be used only with appropriate approval and oversight, for instance by a public health authority, an Institutional Review Board, or a Privacy Committee. Oversight should include assessment of the goals, the information to be used, and the individuals who will have access to the information. The organizations that provide oversight should include representatives from the patient community.*

* *The specific uses of personal medical information should be stated publicly.*

* *The number of individuals who have access to personal medical information should be minimized. All should be trained in, and attest to, requirements for maintaining confidentiality.*

It is essential to recognize that:

* *Some critical questions require use of fully identified information. An example is the need to link to the National Death Index to determine whether a person who received a specific kind of treatment and who isn't still receiving care is still alive. It is also necessary to identify individuals in order to invite them to participate in clinical trials that might be of interest to them.*

* *It is not possible to obtain individual consent for all uses of individuals' data. In many situations, it is essential to evaluate information from individuals whose experience is truly representative of the entire population eligible for treatment. Relying solely on individuals who actively consent provides an unrepresentative picture. Those who consent are typically the ones who are easiest to contact (more likely to remain at the same address, to have telephones, to be available to receive a request, to understand an outreach message), and to be at higher or lower than average risk of the outcomes of concern (either is possible). Evaluating only the people who give consent makes it impossible to know if the results are correct.*

* *It is not possible to notify all individuals personally about all uses of their data. Many important questions that can be answered by use of personal medical information arise years after a person has received care. The individual may have changed care providers or have left the community, and so be effectively untraceable. Intensive follow up methods that might work for a small number of people are impossible for populations that can reach millions.*

* *Opt out provisions, which allow individuals to bar the use of their information, can lead to incorrect answers for everyone, partly because of the difficulties of contacting individuals noted above, but more importantly because the people who opt out might differ in important ways from those who do not. It is particularly important to evaluate the*

experience of people without regard to whether they eventually had either a bad or a good outcome. And it is important to be able to know the proportion of all people who received a treatment who experienced both good and bad outcomes.

2. ***What big data uses will have the greatest impact for making the learning health system successful?***

   Reply: We should expect to gain a much better understanding of the overall safety of medical treatments, of the performance of medical devices like artificial joints, of the outcomes of treatments for conditions like heart disease and diabetes. We have already seen critically important successes. For example, we now know that vaccinating infants and children does not cause autism. Many studies using the records of millions of infants and children were performed to provide convincing evidence of vaccines' safety.

   We will gain a much better appreciation about which treatments work best for which people. We have very limited data on how well and how safely even widely used treatments work for children, the elderly, pregnant women, or people who differ in their genetic makeup.

   We will understand better whether there are important differences between different health care systems in their overall success rates in treating specific conditions.

   We will also be able to provide better community-wide care, by understanding how conditions like asthma, obesity, high blood pressure, and diabetes are, where they are concentrated both in geographic communities and in specific age, gender, or other groups. This understanding will allow targeting of tailored interventions to the communities and individuals who are most in need. It will also be possible to measure the impact of interventions to address these problems.

   We will be able to monitor the occurrence and spread of infectious conditions, like influenza and tuberculosis, to guide timely public health response.

3. ***How should policies to protect privacy and security, and honor ethical obligations, be adjusted for different aspects of the learning health system; specifically, for (1) general data analytics in which an individual's identity is not required, and for (2) precision medicine, in which patient identity is necessary?***

   Reply: As noted above,

   * *Evaluations should use the least amount of information required to answer a question. Often, all that is needed is the number of people treated and the number with the particular outcome of interest.*

   * *Information should be used only with approval and oversight, for instance by public health authority, an Institutional Review Board, or a Privacy Committee. Oversight should include assessment of the goals, the information to be used, and the individuals who will have access*

to the information. The organizations that provide oversight should include representatives from the patient community.

* *The specific uses of personal medical information should be stated publicly.*

* *The number of individuals who have access to personal medical information should be minimized. All should be trained in, and attest to, requirements for maintaining confidentiality.*

4. What challenges to researchers face when they try to aggregate and analyze diverse types of data, some covered and some not by HIPAA? What are potential solutions to these challenges?

Reply: The full value of electronic information rests on the ability to work with detailed data derived from the care of large, representative populations. However, it is not always necessary to share all of the data in order to make effective use of it. Distributed data networks that minimize the need to aggregate individual data are increasingly powerful and should be considered when they are appropriate. These methods move the analyses to the data systems that already possess the data and return results that can be combined across multiple sites. The Food and Drug Administration (FDA), the National Institutes of Health, and the Patient Centered Outcomes Research Institute (PCORI) have created distributed data networks to support some of their needs.

The FDA's Mini-Sentinel program is an example of a large, successful program that uses distributed analysis as its foundational approach to obtaining critical public health information while minimizing the need to aggregate person-level data. The FDA developed the program in response to a Congressional mandate to use electronic health data from over 100 million people to monitor the safety of marketed medical products. Every week the program's coordinating center sends computer programs to 18 separate health plans and insurers to answer FDA's questions in support its public health mission. An example of such a computer program might: 1) evaluate everyone to identify those who started treatment with any of the five specific medicines used to treat high blood pressure and continued treatment for at least two years, 2) among these people, identify everyone who was assigned a first diagnosis of a particular intestinal disorder, and 3) count the number of people identified in 1) and 2) in each of several different age and sex groups. In this example, the only information that the health plan provides to FDA is counts of the number of individuals exposed and the number with the outcome of interest in each age and sex category.

These organizations are able to execute these programs because each has already converted its own patient level data into an identical format. However, each organization maintains both physical and operational control over these transformed data sets. When the organizations receive the programs that address FDA's requests, they execute the programs behind their own firewalls and provide the results to FDA.

The example above, examining intestinal problems complicating treatment with blood pressure medications confirmed FDA's concerns that a particular medication, olmesartan, was associated with a higher risk of an intestinal problem, sprue-like enteropathy, and led the FDA to issue a formal drug safety communication (www.fda.gov/Drugs/DrugSafety/ucm359477.htm). In this example, the experience of hundreds of thousands of people who took the medicines in question was queried without requiring the organizations that already had the data to share any person level information.

In other situations, it is necessary to combine these distributed methods with sharing of information about a fraction of the people involved. An example is the evaluation of the risk of a different kind of intestinal complication following administration to infants of a specific type of rotavirus vaccine. In this example, distributed computer programs identified over one million infants whose records showed they had received the vaccine in question. Among these, several hundred were assigned a diagnosis of intussusceptions, the complication in question. It was necessary for experts to review the records of these several hundred infants in order to apply specific rigorous diagnostic criteria established by an international group that sets standards for evaluations like this. The review confirmed the diagnosis in approximately half of the cases, and also confirmed that the infant had received the specific vaccine in question on the date recorded in the electronic data. This evaluation led to a change in the product's prescribing information. (www.fda.gov/BiologicsBloodVaccines/SafetyAvailability/ucm356758.htm)

It is important to emphasize both the strengths and limitation of distributed analysis methods. At best, they enable sophisticated assessments without making identifiable information available to additional people beyond those who routinely work with it.

It is necessary, though, for these organizations to execute these programs on data that contains identifiers. It is often necessary for these programs to scan the records of all individuals to identify the ones whose information is relevant to a particular question.

In addition, some assessments require some data from some individuals to be accessible to people who don't regularly work with those individuals' data. Some analyses require direct identifiers for some individuals. The example noted above of linking full identifiers to the National Death Index to determine whether individuals who stop receiving health care have died illustrates this need. Generally, the sharing of identifiers is necessary when essential components of a single individual's data are spread across more than one organization. It is often possible to remove direct identifiers from the analysis file after the link is established. However, for many purposes it is important to be able to audit the result; this requires maintaining a crosswalk list that allows re-identification of the individuals so that the original information can be verified in the source location.

When identified information must be shared, it should be stored in protected locations that are accessible only to authorized individuals for approved uses. Data enclaves are one way of storing such data and making it accessible.

5. ***How do systems let patients know that their data might be re-used for learning purposes and that it is a part of a learning health system?***

   Reply: Several approaches should be considered, including informing individuals at the time they receive care, making this information publicly available in the organizations' publications, websites, and other communications.

6. ***What recommendations would you make to help keep policy at pace with or ahead of technology?***

   Reply: Policy should require minimizing the use of identified data consistent with allowing learning activities to occur.