Webinar * May 14, 2013

Privacy and Progress in Whole Genome Sequencing

Recommendation 2: Data Security and Access to Databases

Dixie B. Baker, Ph.D.
Martin, Blanck and Associates

Points of Agreement

1. "Absolute privacy is not possible" (p 6) – parts of our genomic information are clearly visible (e.g., eye color), and we leave a trail of DNA behind us (e.g., saliva on coffee cup)
   - Similarly "absolute security" is not possible, given our dependence on clinical information to provide quality care and to advance medical science
2. Recognition that whole genome sequence data substantially raise the privacy risks (over other medical information) for both the individual and blood relatives (who most likely did not consent)
3. A clear separation between research and clinical contexts is no longer sustainable or desirable
   - Security policy and protective mechanisms need to be consistent, recognizing the reciprocal relationship between these contexts

Points of Agreement

4. "Strong security protections enable individuals to determine autonomously their preferred level of data and information sharing. When individuals have control and can govern sharing of their data at a level with which they are comfortable, they are more likely to have trust in the research or clinical enterprise, and are more likely to participate and share data, benefiting society generally." (p 72)
   - Requires technology to electronically capture and enforce individuals' choices
   - Adherence to Fair Information Practices Principles and comprehensive legal protections are essential
5. "Today's policies must be crafted specifically enough to be actionable and targeted to address our current concerns, yet agile enough to ensure that we do not constrain our ability to adapt to evolving technology, research, and social norms related to privacy and sharing" (p 101)
   - The HIPAA Security Rule's risk-based approach is a good example of such agility

Points Challenged

1. Possession of a whole genome sequence data file by itself affords the individual "practical obscurity" (p 83)
   - "Security by obscurity" was long ago rejected by security engineers and cryptographers, along with the ostrich and the nude emperor

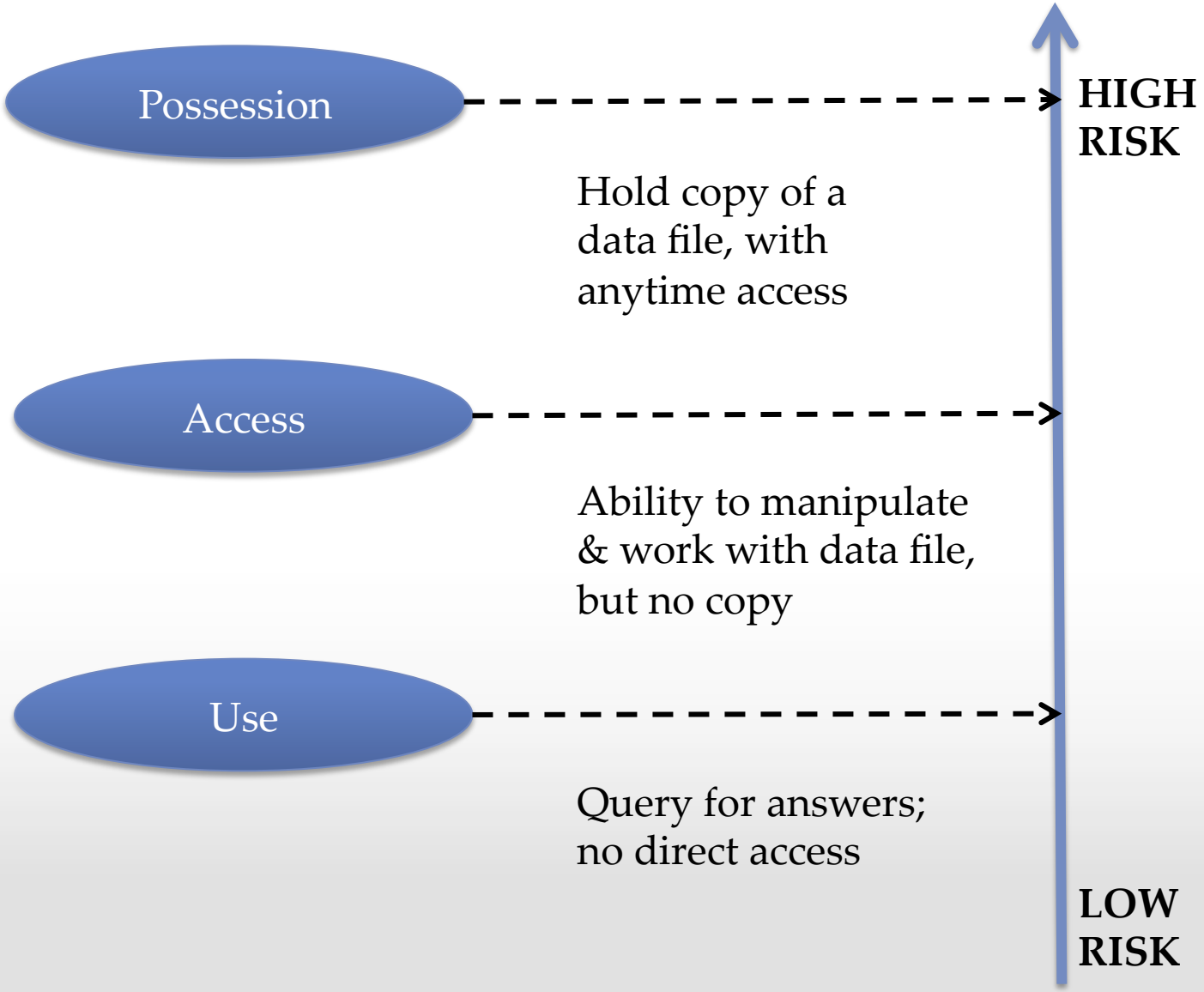
Whole Genome Sequence Data: the Ultimate Identifier

- Whole genome sequence data are inherently unique to the individual, rendering them the ideal "biometric identifier" – one of the 18 data elements of identifiability defined by HIPAA
- A whole genome sequence includes many clues for narrowing the identity possibilities, such as the presence or absence of Y chromosome, visible genetic characteristics (e.g., eye color, hair color, height), and genetic heritage
- Accelerating advances in genetic and "big data" technologies challenge the presumption that any health data can be "de-identified"
  - E.g., Google – and efforts like the "Big Data Research and Development Initiative" (cited on p 85)

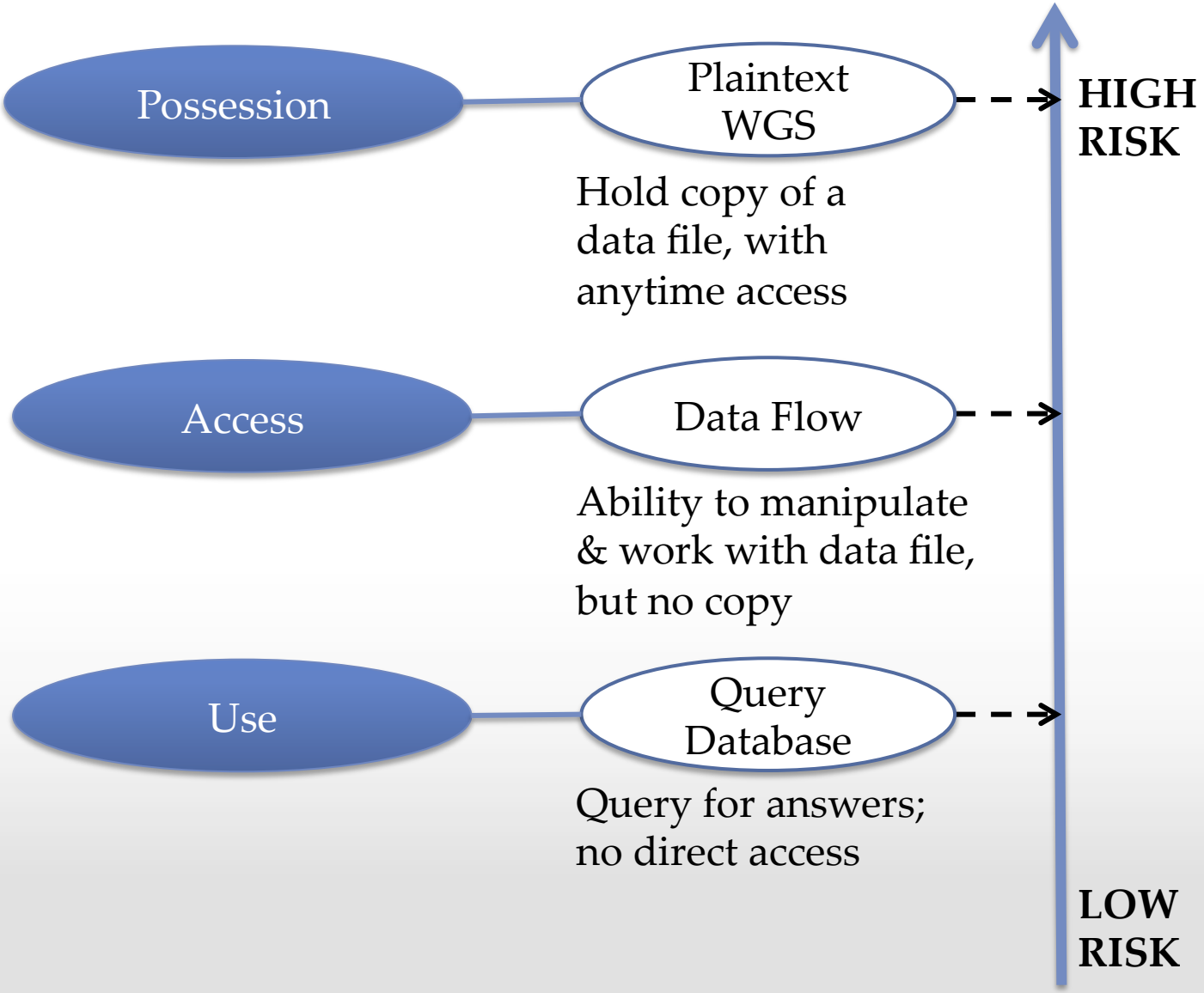
Points Challenged

2. "To determine what baseline privacy protections should be, we need to distinguish between access to, use of, and possession of whole genome sequence data. To *possess* whole genome sequence data is to have a copy of the data file and, therefore, to have access to it at any time. Having *access* to data implies the ability to manipulate and work with the data files. … The *use* of data refers to seeking answers to questions by analyzing the data… without having either access to or possession of the data." (p 47)
   - This characterization may be overly simplistic, and does not accurately convey variables that affect risk

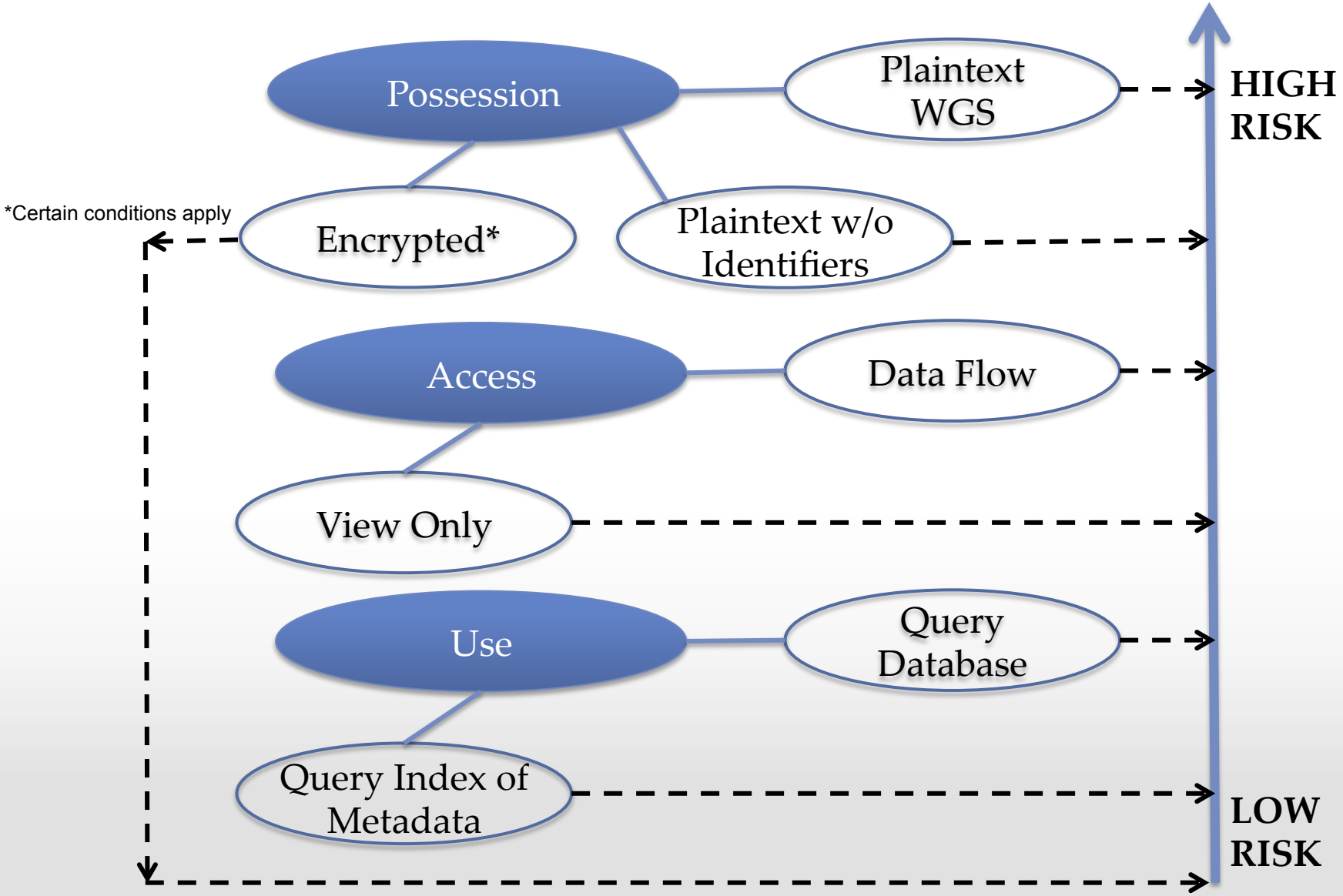
Need for More Granular Distinctions



Need for More Granular Distinctions



Need for More Granular Distinctions



Points Challenged

3. "Complete destruction of whole genome sequence data is likely impossible" because although the primary researchers can destroy their data files, as specified in the consent documents, "the destruction of distributed copies of associated data files may not be feasible as distributed genome sequence data files can be stored on local computers or network servers." (p 94)
   - This is a policy and management issue, not a technology limitation
   - Need to enforce "minimum necessary" rule

Some Final Thoughts

1. HIPAA security provisions should be extended to apply to all entities that handle individually identifiable health information (including whole genome sequences, with or without "traditional identifiers")
2. New legal protections comprehensively addressing the collection, protection, and use of genomic data are needed, and should establish a consistent floor across all states
3. Policy and standards should be developed for incorporating genomic data into electronic health records (EHRs), including whole genome sequences, genetic test results, and clinical observations