Making Genetic Data Public: Privacy Considerations for ClinVar

Megan Martin
Sanjiv Narayan

DATASCI-231
Summer 2023

Agenda

1. Introduction to ClinVar
2. HIPAA privacy risk considerations
3. Intended Uses, Risks and Privacy Analyses
4. Unintended Uses, Risks and Privacy Analyses
5. Re-identification example
6. Conclusions and Recommendations
7. Discussion Questions

Introduction to Clinvar

- Online database of genetic variants and associated clinical symptoms
- Started in 2013 and is used heavily by clinicians, genetic testing companies, and researchers
- Database is freely searchable online and database downloads are available
- Submissions are made by volunteer institutions who are responsible for consent. I.e. ClinVar is a “data broker”.

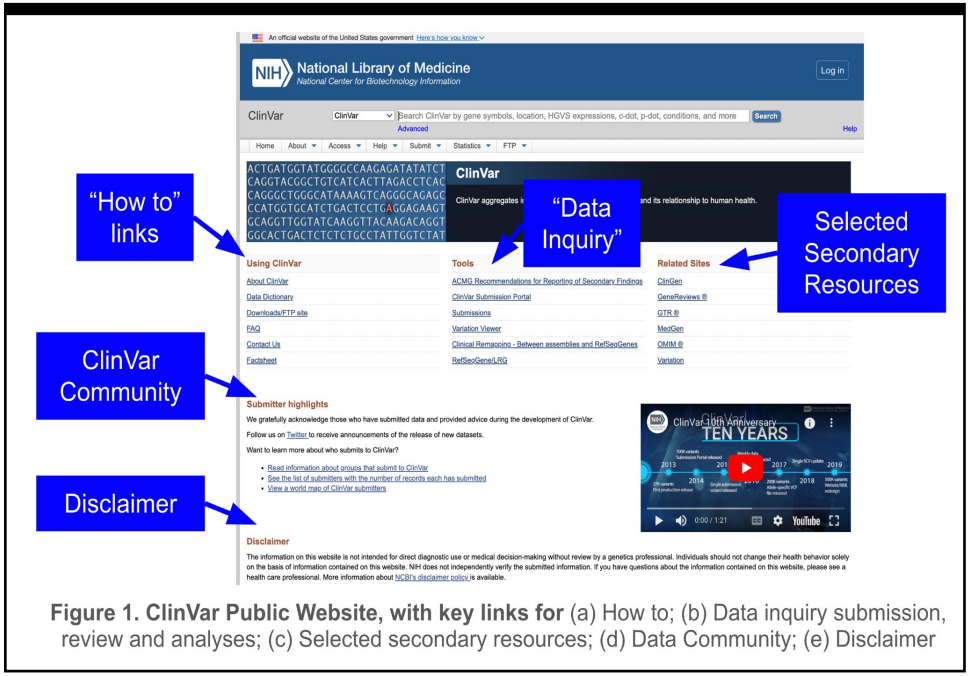


Figure 1. ClinVar Public Website, with key links for (a) How to; (b) Data inquiry submission, review and analyses; (c) Selected secondary resources; (d) Data Community; (e) Disclaimer

How Much Protection does HIPAA Afford?

- Under HIPAA, protected health information (PHI) includes genetic information, and is regulated.
- Clinvar likely qualifies as comprising PHI, as it provides genetic variant information for individuals.
- Data sharing by ClinVar likely falls under 4th provision of HIPAA for the “public health good.”
- Note: HIPAA covers health entities - Not data
  ClinVar isn’t a covered entity. Moreover, many users of publicly available ClinVar data will NOT be bound by HIPAA.

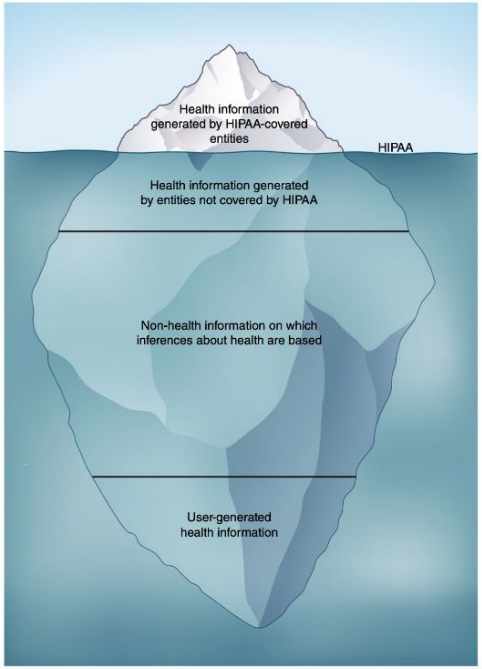


Figure 2. Data covered and not covered by HIPAA

HIPAA is the tip of the iceberg!

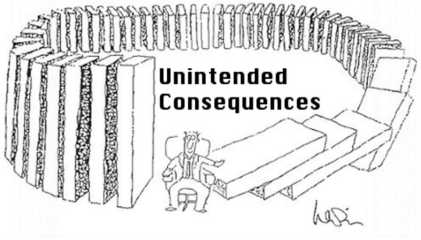
Intended Use

1. Diagnosing and treating a patient who submits data
2. Diagnosing and treating other patients
3. Research and education for above
4. Developing new treatments - including corporate efforts

Risks

1. Privacy - loss of confidentiality, errors in data, processing, dissemination.
2. Belmont: respect for persons (consent to non-HIPAA actors, secondary uses), Benefience (likely met), Justice (inequitable data collection, inequitable benefits due to health care disparities, No compensation for new inventions);
3. Nissenbaum: Corporate uses may violate contextual norms;
4. Mulligan: risks in all 5 dimensions;
5. Solove: surveillance, invasions, decisional interference

Unintended Use



1. You and your family's genetic data for sale (e.g. 23andMe)
2. You and your family's genetic identity aggregated with other data, for sale
3. Re-identification (you and your family)
4. "Infinite" uses over time

Risks

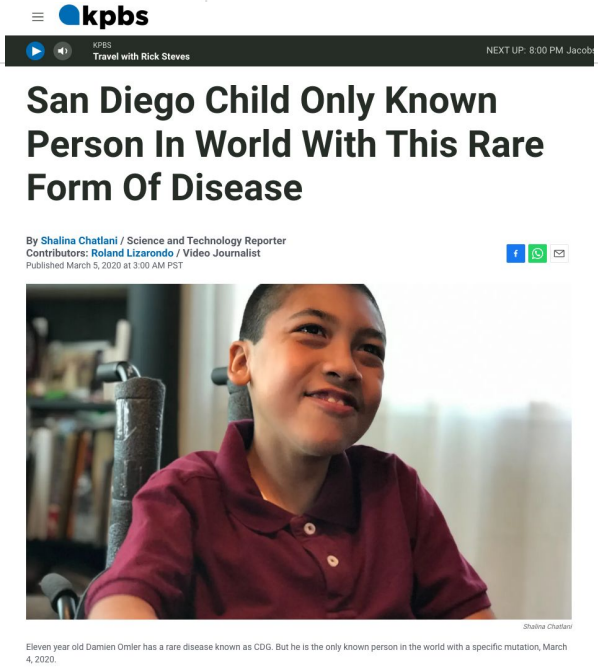
1. Privacy - loss of confidentiality, errors in data, processing, dissemination.
2. Belmont: respect for persons (consent to secondary uses?), Beneficence (unclear public good), Justice (inequitable risks; unclear benefits);
3. Nissenbaum: Data-for-sale, secondary uses may violate contextual norms;
4. Mulligan: risks in all 5 dimensions. From-to: Employers-patient; insurance company-patient, societal stigmatization-patient;
5. Solove: surveillance, invasions, decisional interference

Re-identification risk example

Variation *Location*	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Review status
NM_015949.3(**GET4**):c.837A>G (p.Ile279Met) *GRCh37:* Chr7:933550 *GRCh38:* Chr7:893913	GET4	I279M	**GET4**-related condition	Uncertain significance (Jan 20, 2020)	criteria provided, single submitter

Interpretation (Last evaluated)	Review status (Assertion criteria)	Condition (Inheritance)	Submitter	More information
Uncertain significance (Jan 20, 2020)	criteria provided, single submitter (ACMG Guidelines, 2015) Method: clinical testing	- GET4-related condition (Autosomal recessive inheritance) Affected status: yes Allele origin: paternal	Undiagnosed Diseases Network, NIH Study: Undiagnosed Diseases Network (NIH), UDN Accession: SCV002030284.1 First in ClinVar: Dec 12, 2021 Last updated: Dec 12, 2021	

Comment:
This individual has been published in PMID: 32395830.
Number of individuals with the variant: 1
Clinical Features:
Microcephaly (present) , Abnormal thorax morphology (present) , Abnormal scapula morphology (present) , Abnormal clavicle morphology (present) , Abnormal skull morphology (present) , Abnormal corpus callosum morphology (present) , Abnormal foot morphology (present) , Abnormal cerebral ventricle morphology (present) , Ventriculomegaly (present) , Delayed CNS myelination (present) , Abnormal brainstem morphology (present) , Abnormal cerebral white matter morphology (present) , Coxa valga (present) , Abnormal acetabulum morphology (present) , Abnormality of bone mineral density (present) , Atrophy/Degeneration affecting the brainstem (present) , Corpus callosum atrophy (present) , Abnormal skeletal morphology (present) , Abnormal subarachnoid space morphology (present) , Reduced brain N-acetyl aspartate level by MRS (present) , Cerebral white matter atrophy (present) , Widened cerebral subarachnoid space (present) , Hip subluxation (present) , Abnormal pelvis bone morphology (present) , Enlarged sylvian cistern (present) (less)
Zygosity: 1 Single Heterozygote
Age: 10-19 years
Sex: male
Tissue: blood



Conclusions and Recommendations

- Clinvar serves an essential role in diagnosing patients with genetic diseases. This is key however:
- ClinVar should consider differential privacy practices, particularly controlling access through requiring account set-up
- Clinvar should provide clear protocol for PHI/PII identification and correction, with process for correcting/removing previous static downloads
- Require submitters to confirm proper consent procedures and store consent document on file for full transparency
- Declare risk of identification through secondary data aggregation in intended use section
- The application rises to the level of a full privacy impact assessment

Question 1

- What protections currently exist in the law? How extensive should we review the legal side?

Legal Considerations

- GINA: Genetic Information Non-discrimination Act protects against use of genetic information in employment or health insurance coverage decisions
  - Does not apply to life insurance, long-term insurance, or military personnel/families
- GDPR: unique genetic variants combined with additional data fields in ClinVar is likely not adequately “irreversibly de-identified”
- FTC: genetic information is covered under biometric information
  - Any consumer protections would relate to sharing practices and terms of use between patient and genetic testing institution
- But, it is not clear who would enforce these..

Question 2

- Should genetic testing companies be heavily regulated and held to international standards?

Question 3

- For contextual integrity, should we bring in relevant literature regarding patient perspectives on research participation?

References

- Callaway, E. Supercharged crime-scene DNA analysis sparks privacy concerns. Nature 2018 Vol. 562 Issue 7727 Pages 315-316 https://www.ncbi.nlm.nih.gov/pubmed/30327515
- L. J. Dewar, M. Alcaide, D. Fornika, L. D'Amato, S. Shafaatalab, C. M. Stevens, et al. Investigating the Genetic Causes of Sudden Unexpected Death in Children Through Targeted Next-Generation Sequencing Analysis. Circ Cardiovasc Genet 2017 Vol. 10 Issue 4 https://www.ncbi.nlm.nih.gov/pubmed/28807990
- Harris Interactive. 2001. Misconceptions and lack of awareness greatly reduce recruitment for cancer clinical trials. *Health Care News* 1(3).
- Kuo TT, X. Jiang, H. Tang, X. Wang, A. Harmanci, M. Kim, et al. The evolving privacy and security concerns for genomic data analysis and sharing as observed from the iDASH competition J Am Med Inform Assoc 2022 Vol. 29 Issue 12 Pages 2182-2190 https://www.ncbi.nlm.nih.gov/pubmed/36164820
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., & Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research*, *44*(D1), D862–D868. https://doi.org/10.1093/nar/gkv1222 https://doi.org/10.1093/nar/gkx1153
- Mulligan, Deirdre K., Koopman, Colin and Doty, Nick (2016). Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. Philosophical Transactions of The Royal Society A: Mathematical Physical and Engineering Sciences, 374(2083):20160118 (December 2016). http://doi.org/10.1098/rsta.2016.0118
- Nissenbaum, Helen F. (2011). A Contextual Approach to Privacy Online. Daedalus 140:4 (Fall 2011), 32-48. https://ssrn.com/abstract=2567042
- Shabani M, Dyke SOM, Marelli L, Borry P. Variant data sharing by clinical laboratories through public databases: consent, privacy and further contact for research policies Genet Med 2019 Vol. 21 Issue 5 Pages 1031-1037 https://www.ncbi.nlm.nih.gov/pubmed/30293992
- Shabani, M., & Borry, P. (2018). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European journal of human genetics : EJHG*, *26*(2), 149–156. https://doi.org/10.1038/s41431-017-0045-7
- Solove, Daniel J. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154:3 (January 2006). https://ssrn.com/abstract=667622

Data Submission Process

- Volunteer submitters from genetic testing companies, research centers, patient registries, and medical geneticists
- Clinvar assumes submitter has obtained appropriate consent
- Review of consent documents from top 3 submitters did not have options to opt-out of database sharing
- Required and optional data fields could reveal date of testing, clinical features, specific age, testing laboratory, and ethnicity/geographic origin