

GRDR Working group on Informed Consent for Rare Disease Registries Linked to Biorepositories

Public access to de-identified data of patients with rare diseases: overcoming challenges, problems and risks

Clement J. McDonald, M.D.

Director, Lister Hill National Center for Biomedical Communications

U.S. National Library of Medicine, NIH, HHS

December 13, 2010
Rockville, Maryland



DE-IDENTIFICATION

December 13, 2010



Definition

- **Specified in HIPAA regulations**
- **18 specific identifiers**
 - Includes all expected facial identifiers (e.g., SS#, name)
 - Full dates
 - Year of birth OK up till age 90
 - Details in

Not the same as anonymization

- The regulations accept that it is close but not the same
- But.... There are traps that could ensnare the best efforts
- For anonymization ,a good way to think of the goal is to avoid the situation where:
 - The information would tell you more about the patient than you already know.
- Some who worry say –what if you knew all of the patient’s laboratory tests, you could then uniquely identify the patient and find his/her record, but...



Different kinds of data present different problems

- Structured data base data— better than narrative
 - You can remove the fields known to contain HIPAA identifiers
 - e.g. Name, hospital number
 - You can apply good judgment to the other variables and decide whether to release or not—because you know what they carry
 - However, always check the data to be sure free text has not slipped into a field meant to be a numerically valued or Likert scale



Narrative content

- Can be scrubbed (by hand) or by software for names and other identifiers
- Commercial software products exist (DeID) – as do researcher products (see paper published in Biomed Central)
- But trickier than fully-structured data.

HIPAA de-identification has been criticized unfairly

- Some critics assert that is easy to re-identify
- Based on a report (on web: **A Technical and Legal Study** by Salvador Ochoa et al. May 6, 2001- from MIT) that describes success at re-identifying a data set
- But that data set did NOT follow HIPAA rules- and the authors assert they could NOT re-identify when HIPAA rules were applied.

Many de-identified data sets have operated without problems for years

- E.g. – NCI's SEER cancer registry data
- Data sets with relatively few fields tend to be safe-ish
- Statistical content tends to be safe...
- But not guaranteed to be so

There are challenges

- It is possible in some circumstances to reconstruct patient level data from marginal statistics
- A well-publicized paper in last 2 years showed the possibility of using statistics re: DNA variations
- dbGAP from NLM's NCBI does not report some statistics because of this reality

Permission makes a difference

- PatientsLikeMe.com
- web site that collects very detailed data from patients who share it with each other, AND provide it in statistical form for public viewing
- The founder has a very interesting premise about the obligations of individuals to share their data to speed the arrival of cures
- He is compelling

Patients Like Me Web Site

PatientsLikeMe : ALS/Motor Neuron Disease Community - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.patientslikeme.com/als/community

Most Visited Getting Started Latest Headlines

PatientsLikeMe : ALS/Motor Neuron ...

patientslikeme®

Share your experiences, learn from others, and connect with patients just like you.

Join Now! (It's free!)

Already a member? Log in

Patients | Treatments | Symptoms | Research

Select a community **ALS** including PLS and PMA

Home > ALS Community

ShareThis

The PatientsLikeMe ALS/Motor Neuron Disease Community



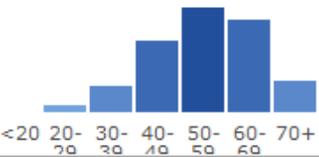
Find Patients Just Like You >>

Join Now! (It's free!)

ALS, or amyotrophic lateral sclerosis, is a neurodegenerative disease caused by the degeneration of motor neurons. Around 3,000 to 5,000 new cases of ALS are diagnosed every year in the United States alone. Imagine knowing what treatments work for people just like you – and having the ability to easily connect with those people. That's what makes PatientsLikeMe different. By sharing your symptoms, treatments and more, you can gain insight into what affects your condition while helping others learn from your experiences. Together, we can achieve more. [Begin making a difference today!](#)

Who Will You Find?

4,948 total patients
154 Profiles updated this week
57 New patients this month



Age Group	Number of Patients
<20	~10
20-29	~20
30-39	~40
40-49	~60
50-59	~80
60-69	~70
70+	~30

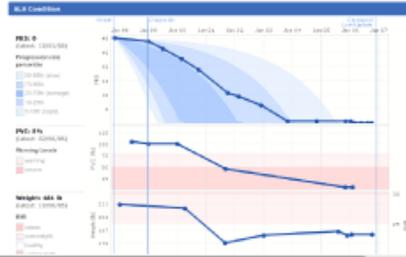
42% Females



What Can You Learn?

Your Health Profile

By creating and sharing your health profile with the community, you can learn more about how your condition affects you and help others learn from your experience.



Read i2.ytimg.com

Special problem with very rare diseases

- If a disorder is very rare on the order of 100's then the inclusion of that disease identifier in the patient's records could provide new private information to those who know he/she has the disorder..
- Realize that the fact the patient has the disorder is not news, it's the other things that may be in the database.



Bio Bases

- Presume that DNA (at least) will be carried in the database.
- Specimens were regarded as not PHI in the original HPAA regulations, as long as they were not associated with any identifiers.
- That has been tightened up.

What to do

- De-identify
- But either make nothing (or only very limited and privacy-safe content) *publically* available.
- What I mean by publically is that anyone in the world can look at it
- Create some safety barrier
 - Signed agreements from researchers
- If there is interest from affected families in looking at the data-might limit that look to statistics and also get an agreement that they won't try to de-identify



Consents

- You will likely require consents under almost any circumstance to make reasonable scientific use of the biobank
- So also write into the consent-step wise approvals for research, with separation among:
 - Permission to contact
 - Permission to access all of the data – but no contact
 - Permission to access de-identified data
 - Permission to access the specimen and/or specimen data



GWAS data

- Includes all of the patient observations AND either genetic data and/or access to tissues.
- NIH developed protocol for handling requests to such data
- In short, the institute that provided the GWAS data manages and approves requests for access
- For details see:
http://privacyruleandresearch.nih.gov/pr_08.asp

