

HHS Public Access

Author manuscript *Neuroinformatics.* Author manuscript; available in PMC 2017 December 06.

Published in final edited form as: Neuroinformatics. 2015 January ; 13(1): 1–6. doi:10.1007/s12021-014-9248-z.

Genetic Data Sharing and Privacy

Marco D. Sorani, John K. Yue, Sourabh Sharma, Geoffrey T. Manley, Adam R. Ferguson, The TRACK TBI Investigators, Shelly R. Cooper, Kristen Dams-O'Connor, Wayne A. Gordon, Hester F. Lingsma, Andrew I. R. Maas, David K. Menon, Diane J. Morabito, Pratik Mukherjee, David O. Okonkwo, Ava M. Puccio, Alex B. Valadka, and Esther L. Yuh Department of Neurological Surgery, University of California, San Francisco, 1001 Potrero Avenue, Bldg 1, Room 101, San Francisco, CA 94110-0899, USA

Keywords

Data; Privacy; Genetic; Biospecimen; Biobank; Sharing

Capitalizing on the Promise of Genetic Data

Genetic data has provided valuable insights into disease cause and risk as well as drug discovery and development in neuroscience. For example, human genetics studies have provided insights into cognition (Glahn et al. 2013) and psychiatric disorders (Kao et al. 2010). The genetic basis of several inherited disorders such as Down's Syndrome and Tay-Sachs disease are well known, and other associations such as the role of APOE in Alzheimer's disease are still extensively studied. However, despite advances in understanding the human genome, there are concerns about the privacy of genetic data and potential discrimination resulting from its disclosure, and there has been incomplete oversight of genetic testing (Scheuner et al. 2008).

At the same time, there have been increased efforts to share research data to enable scientific discovery and achieve cost efficiencies. It has become clear that no scientist can guarantee absolute privacy, and it is also increasingly recognized that research will work better if scientists have more information about the people they study and that being identifiable has some benefits (Angrist 2013). There are examples of pioneering efforts in neuroscience research. The fMRI Data Center is a leader in open-access data sharing in the functional neuroimaging community, overcoming logistical, cultural and funding barriers (Mennes et al. 2013). Similarly, the INCF Task Force on Neuroimaging Datasharing has started work on tools to ease and automate sharing of raw, processed, and derived neuroimaging data and metadata (Poline et al. 2012).

In the United States, legislation such as the Health Insurance Portability and Accountability Act (HIPAA) (Gostin 2001) and the Genetic Information Nondiscrimination Act have attempted to limit access to sensitive data and discrimination related to health insurance and

Correspondence to: Marco D. Sorani.

Conflicts of Interest The authors declare that they have no conflicts of interest.

employment, but it has been known for over a decade that seemingly anonymized data can be related to publicly available information to identify specific individuals (Braun et al. 2009) using diagnosis codes (Tamersoy et al. 2010), rare visible disorders (Eguale et al. 2005), allele frequencies (Craig et al. 2011), place and date of birth (Acquisti and Gross 2009), a combination of a surname with age and state (Gymrek et al. 2013), and patient health location visit patterns (Malin 2007). Re-identification methods have included genotype-phenotype inferences, family structures, and dictionary attacks (Malin 2005).

In total, these facts have changed the goals of many research organizations from making data re-identification impossible to making it highly improbable and educating stakeholders about the issues and risks, while enhancing research collaborations by sharing data. Here, we discuss data privacy and sharing approaches, we provide recommendations and describe our own experiences in the context of biobanking, and we look ahead to address challenges and opportunities for data privacy and sharing.

Data Sharing "Carrots" and "Sticks"

Data sharing is often driven by a set of incentives and consequences. Benefits include a desire to "democratize" data, an evolution towards more "big science" collaboration, a desire to minimize the burden on research participants, technical developments such as web based databases, standardized data sharing guidelines, and opportunities for exclusivity in manuscript submission and citation of data sets. For example, the "data paper" allows researchers to publish their datasets as a citable scientific publication, gives credit that is recognizable within the scientific community, and ensures the quality of the published data and metadata through the peer review process (Gorgolewski et al. 2013). Consequences and risks of not sharing data include declining financial resources, a need to manage data beyond the lifecycle of a grant, and requirements by journal editors. Challenges to effective sharing remain, including the removal of disincentives for data sharing by industry, the reduction of litigation risks, the availability of patient level data, and the willingness to foster discussion in cases of differences of interpretation of data.

Privacy Strategies, Processes and Technologies

To address privacy objectives, a wide variety of strategies have been proposed to protect sensitive data. Many organizations rely on the Safe Harbor Standard of the HIPAA Privacy Rule, which enumerates 18 identifiers that must be suppressed (Malin et al. 2011), though HIPAA has also been criticized by the Institute of Medicine for provisions that seem to hinder data access while failing to provide substantive privacy protection (Franc et al. 2011). De-identification procedures have been applied to a variety of types of free-text data, including electronic health records (Meystre et al. 2010), laboratory and clinical narrative reports (Friedlin and McDonald 2008), discharge and order summaries (Aberdeen et al. 2010), and pathology reports (Thomas et al. 2002), as well as non-text data formats such as images (Gonzalez et al. 2010) and geocode data (Cassa et al. 2006). Johnson et al. tested a method for generating global unique identifiers to link data and specimens by sending encrypted information to a server application with over 8000 individuals in an autism study.

They implement a balance between distinguishing individuals to gain research insights and protecting confidentiality (Johnson et al. 2010).

A recent survey of biopharma companies found that the most common data coding practice was de-identification. Only 10 % of companies anonymized their data. Most reported retaining a secondary key internally either by a party independent of the one responsible for de-identification or by the same party but with restricted access (38 %). De-identification programs are difficult to implement, cumbersome, costly, inefficient and offer little added privacy protection (Franc et al. 2011).

Privacy processes and technologies have evolved together. Processes include establishing an "honest broker" that provides investigators with de-identified or limited datasets under stipulations contained in a signed data use agreement (Liu et al. 2009), and establishing data enclaves where investigators apply to obtain restricted access to data for a limited time with the understanding they will be monitored (Rodgers and Nolte 2006). Some of these strategies may be difficult to implement and explain to patients and other stakeholders, and they may not provide added benefit proportional to the cost. At the same time, many automated algorithms for anonymizing data are now available (Loukides et al. 2010). Algorithms include family relation linking (Malin 2006), encryption (Landi and Rao 2003), and hiding functions (Huang et al. 2010). In addition, database software systems are a mature, increasingly ubiquitous technology and come with robust security and audit functionality. For example, a recent public epidemiology project implemented a database system for data privacy using a multi-layered role and right-of-access control plus de-identification (Meyer et al. 2012). Researchers, particularly in academia, must be prepared to move past spreadsheets.

Biobanks as Laboratories for Data Sharing

The emergence of genomic technologies has spurred rapid growth in the collection of biosamples and the development of biobanks. While some types of research data can be useful even if they are "permanently" anonymized, biosample data poses unique challenges for security from a research perspective. Genetic data must be kept private as stipulated in informed consent agreements, but other related data does not face this requirement or must in fact remain identifiable. To achieve patient confidentiality, DNA samples can be relabeled with unique identifiers that are different than the identifiers initially assigned in a clinical trial. This relabeling or double-coding process is referred to as de-identification. With data de-identification, the data belonging to an individual in the clinical environment can still be associated with the same individual in a de-identified research context. An increasingly common research objective involves building analytical databases using de-identified clinical data while enabling the data set to be updated with new pseudonymous data over time (Noumeir et al. 2007).

In the United States, biobanks in academia, government, and industry have implemented a range of sharing and security practices. At Vanderbilt University, a DNA biobank linked to data from an electronic medical record (EMR) system implemented a de-identified mirror image of the EMR, a policy of extracting DNA from discarded blood samples, revision of

standard consent, and procedures for de-identification. About 700 to 900 samples were added per week (Roden et al. 2008). The National Mesothelioma Virtual Bank is based on the caTISSUE Clinical Annotation Engine developed in cooperation with the Cancer Biomedical Informatics Grid and includes paraffin embedded tissues, tissue microarrays, serum and genomic DNA. It provides real-time access to de-identified data depending on user authorization (Amin et al. 2008). Internationally, practices also vary. The Genome Austria Tissue Bank developed data protection tools and considered ethical, legal and social issues as it changed from a population-based tissue bank to a disease-focused biobank (Asslaber et al. 2007). The National Cancer Center Hospital in Tokyo established a biorepository in 2002 for both de-identified and non-de-identified post-clinical test samples. A portion of samples are transferred to new tubes before and after being frozen. This transfer is the only de-identification procedure (Furuta et al. 2011).

Recommendations and Examples

For those new to multicenter studies and 'big data' analyses of genomic and clinical data, several guiding principles may be useful. Other resources such as the NIH's "Data Sharing Policy and Implementation Guidance" and the NSF's "Data Sharing Policy" should also be consulted.

- Institutional review board (IRB) approval of study designs, informed patient consent of study participation, and understanding of HIPAA requirements are essential.
- A flexible and powerful computing platform for data management is critical but does not have to be complex or expensive. Emailing spreadsheets will lead to problems setting up an open source database with web access has become increasingly feasible.
- There are generally four categories of coding for data security: identified, coded (including single-coded and double-coded or "de-identified"), anonymized, and anonymous. It is necessary to select the appropriate level of data coding to balance patient security and research purposes.
- Providing data alone is not enough to enable effective sharing data dictionaries describing the data and, of course, human support are critical. Common analytical methods and tools further support sharing of metadata and results.
- Realistic expectations regarding sharing (e.g., database updates and speed, publication plans, etc.) and privacy (e.g., who will have access to what information, etc.) will establish trust.
- Thinking ahead to future requirements (e.g., will investigators want to obtain consent to re-contact subjects for follow-up?) and challenges (e.g., is there a plan in case of data security breach?) will enhance likelihood of overall success and sustainability.

In general, collaboration with a cross-functional team of clinicians, statisticians, geneticists, informaticians, and other relevant subject matter experts will also help identify opportunities and challenges.

An Example of Multi-Dimensional Data Sharing: TRACK-TBI

Data-sharing for precision-medicine in neuroscience will require co-mingling of biorepository, brain imaging and functional data, raising specific challenges. Our research group is participating in an example of successful de-identified data sharing: the <u>Transforming Research and Clinical Knowledge for Traumatic Brain Injury (TRACK-TBI)</u>, a series of two large-scale prospective multicenter observational trials for improving traumatic brain injury diagnosis and therapeutic targeting (ClinicalTrials.gov Identifiers NCT01565551, pilot 2010–13; NCT02119182, ongoing 2014–18) (Yue et al. 2013). The pilot phase consisted of three centers and collected data on 599 patients. The TRACK-TBI project applies the official NIH/NINDS TBI Common Data Elements (TBI-CDEs) and standardized collection protocols for biospecimens (Diaz-Arrastia et al. 2014), imaging (Yuh et al. 2013; 2014), and neurocognitive and neuropsychiatric outcome metrics (Dams-O'Connor et al. 2013; Lingsma et al. 2014).

Multicenter patient data, including protected health information (PHI) as defined by HIPAA, is collected under informed consent into a central, custom-designed repository (QuesGen Systems, Inc., Burlingame, CA) that assigns a globally unique identifier. The system manages permissions ranging from 'no-PHI', to 'local PHI only', to 'full-access'. The 'full-access' view is only available to a small quality control and assurance team. Investigators have access to only local PHI from their enrolling center, and only if access is pre-approved by the local institutional review board. Multicenter data are broadly searchable in the no-PHI view by the full group of TRACK-TBI investigators. Ultimately, portions of the 'no-PHI' view of TRACK-TBI data will be publicly accessible for research purposes through the US Federal Interagency TBI Research (FITBIR) informatics system (https://fitbir.nih.gov). Additionally, data use agreements guide the handling of data, and subject timeline data are recoded from dates to days from baseline. These processes have been expanded upon in the current TRACK-TBI study funded for 11 centers with goals of capturing data on over 3000 patients in the next 4 years.

Privacy Concerns and Use

Patients and research subjects have long expressed concern about privacy of health information. In a study by King et al. (2012), participants preferred to be asked for permission before their health information was used for any purpose other than medical treatment (92 %), and they wanted to know the details of the research before allowing the use of their health records (83 %). The study showed that there are some particularly sensitive issues, including family medical history, genetic disorders, mental illness, drug or alcohol related incidents, lists of previous procedures, and current medications. There are also ethical doubts about the ability of cognitively impaired subjects to give informed consent or addicted subjects to participate in studies that involve the administration of drugs of dependence. Tests to identify addicts or predict risk of addiction will raise concerns about

Privacy around genetic data has also become a concern among researchers. Lapses in data security can result in undesired publicity and expense (Benitez and Malin 2010). Kho et al. (2009) write that "to ensure that legislation on privacy does not unduly bias observational studies using medical records, thoughtful decision making by research ethics boards on the need for mandatory consent is necessary". Again, attitudes and practices vary widely. In interviews with administrators at Canadian universities, 47 % of sites required individual patient consent for studies to proceed, but 45 % did not require consent or suggested a notification and opt-out process (Willison et al. 2008). Lane and Schur (2010) proposed that a guiding research principle should be "to generate released data that are as close to the maximum acceptable risk as possible. HIPAA and other privacy measures can perhaps be seen as having had the effect of lowering the 'maximum acceptable risk' level and rendering some data unreleasable."

Going Forward

The ability to integrate clinical, genetic, imaging, and other types of biomedical data will be of tremendous value in ongoing efforts to discover and develop biomarkers and drugs to address unmet medical needs. New technologies and processes can simultaneously support data privacy, data sharing, and research objectives. Here, we described several basic principles as well as experiences from our research group to illustrate that data privacy and sharing can be accomplished together. Our experience highlights an approach in the context of an ambitious basic and clinical research collaboration.

Van Horn and Ball (2008) write that although it will not be "a pain-free process, with increased data availability, scientists from multiple fields can enjoy greater opportunity for novel discoveries about the brain in health and disease." Despite the rapid pace of technological change, it will be critical to think carefully about balancing concerns and objectives. Neuroscientists and neurologists, as well as regulators and patients, can seek opportunities to collaborate with research and clinical colleagues as they advocate for data privacy and high quality science and medicine.

Acknowledgments

Funding Authors supported by Department of Defense United States Army Medical Research Acquisition Activity W81XWH-13-1-0441 (GTM), and NIH Grants NS069409 (GTM) and NS069409-02S1 (GTM), NS067092 (ARF), NS069537 (ARF), Craig H. Neilsen foundation Grant 224308 (ARF) and Wings for Life Foundation Grant WFL-US-008/12 (ARF).

References

- Aberdeen J, Bayer S, Yeniterzi R, Wellner B, Hirschman L, Malin B. The MITRE Identification Scrubber Toolkit: design, training, and assessment. International Journal of Medical Informatics. 2010; 79(12):849–859. [PubMed: 20951082]
- Acquisti A, Gross R. Predicting Social Security numbers from public data. Proceedings of the National Academy of Science. 2009; 106(27):10975–10980.

- Amin W, Parwani AV, Schmandt L, Mohanty SK, Farhat G, Pople AK, Winters SB, Whelan NB, Schneider AM, Milnes JT, Valdivieso FA, Feldman M, Pass HI, Dhir R, Melamed J, Becich MJ. National Mesothelioma Virtual Bank: a standard based biospecimen and clinical data resource to enhance translational research. BMC Cancer. 2008; 8:236. [PubMed: 18700971]
- Angrist M. Genetic privacy needs a more nuanced approach. Nature. 2013; 494(7435):7. [PubMed: 23389508]
- Asslaber M, Abuja PM, Stark K, Eder J, Gottweis H, Trauner M, Samonigg H, Mischinger HJ, Schippinger W, Berghold A, Denk H, Zatloukal K. The Genome Austria Tissue Bank (GATiB). Pathobiology. 2007; 74(4):251–258. [PubMed: 17709968]
- Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. Journal of the American Medical Informatics Association. 2010; 17(2):169–177. [PubMed: 20190059]
- Braun R, Rowe W, Schaefer C, Zhang J, Buetow K. Needles in the haystack: identifying individuals present in pooled genomic data. PLoS Genetics. 2009; 5(10):e1000668. [PubMed: 19798441]
- Cassa CA, Grannis SJ, Overhage JM, Mandl KM. A context-sensitive approach to anonymizing spatial surveillance data: impact on outbreak detection. Journal of the American Medical Informatics Association. 2006; 13:160–165. [PubMed: 16357353]
- Craig DW, Goor RM, Wang Z, Paschall J, Ostell J, Feolo M, Sherry ST, Manolio TA. Assessing and managing risk when sharing aggregate genetic variant data. Nature Reviews Genetics. 2011; 12(10): 730–736.
- Dams-O'Connor K, Spielman L, Singh A, Gordon WA, Lingsma HF, Maas AI, Manley GT, Mukherjee P, Okonkwo DO, Puccio AM, Schnyer DM, Valadka AB, Yue JK, Yuh EL, Investigators T. R. A. C. K.-T. B. I. The impact of previous traumatic brain injury on health and functioning: a TRACK-TBI study. Journal of Neurotrauma. 2013; 30(24):2014–2020. [PubMed: 23924069]
- Diaz-Arrastia R, Wang KK, Papa L, Sorani MD, Yue JK, Puccio AM, McMahon PJ, Inoue T, Yuh EL, Lingsma HF, Maas AI, Valadka AB, Okonkwo DO, Manley GT, Investigators T. R. A. C. K.-T. B. I. Acute biomarkers of traumatic brain injury: relationship between plasma levels of ubiquitin Cterminal hydrolase-L1 and glial fibrillary acidic protein. Journal of Neurotrauma. 2014; 31(1):19– 25. [PubMed: 23865516]
- Eguale T, Bartlett G, Tamblyn R. Rare visible disorders/diseases as individually identifiable health information. AMIA Annual Symposium Proceedings. 2005; 947
- Franc MA, Cohen N, Warner AW, Shaw PM, Groenen P, Snapir A, Industry Pharmacogenomics Working Group. Coding of DNA Samples and Data in the Pharmaceutical Industry: Current Practices and Future Directions-Perspective of the I-PWG. Clinical Pharmacology & Therapeutics. 2011; 89(4):537–545. [PubMed: 21346752]
- Friedlin FJ, McDonald CJ. A software tool for removing patient identifying information from clinical documents. Journal of the American Medical Informatics Association. 2008; 15(5):601–610. [PubMed: 18579831]
- Furuta K, Yokozawa K, Takada T, Fujiwara Y. De-identification Procedure and Sample Quality of the Post-clinical Test Samples at the Bio-repository of the National Cancer Center Hospital (NCCH) in Tokyo. Japanese Journal of Clinical Oncology. 2011; 41(2):295–298. [PubMed: 20852301]
- Glahn DC, Kent JW Jr, Sprooten E, Diego VP, Winkler AM, Curran JE, McKay DR, Knowles EE, Carless MA, Goring HH, Dyer TD, Olvera RL, Fox PT, Almasy L, Charlesworth J, Kochunov P, Duggirala R, Blangero J. Genetic basis of neurocognitive decline and reduced white-matter integrity in normal human brain aging. Proceedings of the National Academy of Science. 2013; 110(47):19006–19011.
- Gonzalez DR, Carpenter T, van Hemert JI, Wardlaw J. An open source toolkit for medical imaging deidentification. European Radiology. 2010; 20(8):1896–1904. [PubMed: 20204640]
- Gorgolewski KJ, Margulies DS, Milham MP. Making data sharing count: a publication-based solution. Frontiers in Neuroscience. 2013; 7:9. [PubMed: 23390412]
- Gostin LO. National health information privacy: regulations under the Health Insurance Portability and Accountability Act. JAMA. 2001; 285(23):3015–3021. [PubMed: 11410101]
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013; 339(6117):321–324. [PubMed: 23329047]

- Hall W, Carter L, Morley KI. Neuroscience research on the addictions: a prospectus for future ethical and policy analysis. Addictive Behavior. 2004; 29(7):1481–1495.
- Huang LC, Chu HC, Lien CY, Hsiao CH, Kao T. Embedding a hiding function in a portable electronic health record for privacy preservation. Journal of Medical Systems. 2010; 34(3):313–320. [PubMed: 20503616]
- Johnson SB, Whitney G, McAuliffe M, Wang H, McCreedy E, Rozenblit L, Evans CC. Using global unique identifiers to link autism collections. Journal of the American Medical Informatics Association. 2010; 17(6):689–695. [PubMed: 20962132]
- Kao WT, Wang Y, Kleinman JE, Lipska BK, Hyde TM, Weinberger DR, Law AJ. Common genetic variation in Neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain. Proceedings of the National Academy of Science. 2010; 107(35):15619–15624.
- Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. BMJ. 2009; 338:b866. [PubMed: 19282440]
- King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. International Journal of Medical Informatics. 2012; 81(4):279–289. [PubMed: 22306206]
- Landi W, Rao RB. Secure de-identification and re-identification. AMIA Annual Symposium Proceedings. 2003:905. [PubMed: 14728411]
- Lane J, Schur C. Balancing access to health data and privacy: a review of the issues and approaches for the future. Health Services Research. 2010; 45(5):1456–1467. [PubMed: 21054366]
- Lingsma H, Yue JK, Maas A, Steyerberg EW, Manley GT, Cooper SR, Dams-O'Connor K, Menon D, Mukherjee P, Okonkwo DO, Puccio AM, Schnyer DM, Valadka A, Vassar MJ, Yuh EL. Outcome Prediction After Mild and Complicated Mild Traumatic Brain Injury: External Validation of Existing Models and Identification of New Predictors Using the TRACK-TBI Pilot Study. Journal of Neurotrauma. 2014 Jul 15. Epub ahead of print.
- Liu J, Erdal S, Silvey SA, Ding J, Riedel JD, Marsh CB, Kamal J. Toward a fully de-identified biomedical information warehouse. (2009). AMIA Annual Symposium Proceedings. 2009:370– 374. [PubMed: 20351882]
- Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. Proceedings of the National Academy of Science. 2010; 107(17):7898–7903.
- Malin BA. An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. Journal of the American Medical Informatics Association. 2005; 12(1):28–34. [PubMed: 15492030]
- Malin B. Re-identification of familial database records. AMIA Annual Symposium Proceedings. 2006:524–528. [PubMed: 17238396]
- Malin B. A computational model to protect patient data from location-based re-identification. Artificial Intelligence in Medicine. 2007; 40(3):223–239. [PubMed: 17544262]
- Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. Journal of the American Medical Informatics Association. 2011; 18(1):3–10. [PubMed: 21169618]
- Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. NeuroImage. 2013; 82:683–691. [PubMed: 23123682]
- Meyer J, Ostrzinski S, Fredrich D, Havemann C, Krafczyk J, Hoffmann W. Efficient data management in a large-scale epidemiology research project. Computer Methods and Programs in Biomedicine. 2012; 107(3):425–435. [PubMed: 21256617]
- Meystre SM, Friedlin FJ, South BR, Shen S, Samore MH. Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Medical Research Methodology. 2010; 10:70. [PubMed: 20678228]
- Noumeir R, Lemay A, Lina JM. Pseudonymization of radiology data for research purposes. Journal of Digital Imaging. 2007; 20(3):284–295. [PubMed: 17191099]

- Poline JB, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, Haselgrove C, Helmer KG, Keator DB, Marcus DS, Poldrack RA, Schwartz Y, Ashburner J, Kennedy DN. Data sharing in neuroimaging research. Frontiers in Neuroinformatics. 2012; 6:9. [PubMed: 22493576]
- Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clinical Pharmacology & Therapeutics. 2008; 84(3):362–369. [PubMed: 18500243]
- Rodgers W, Nolte M. Solving problems of disclosure risk in an academic setting: using a combination of restricted data and restricted access methods. Journal of Empirical Research on Human Research Ethics. 2006; 1(3):85–98. [PubMed: 19385825]
- Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. JAMA. 2008; 299(11):1320–1334. [PubMed: 18349093]
- Tamersoy A, Loukides G, Denny JC, Malin B. Anonymization of administrative billing codes with repeated diagnoses through censoring. AMIA Annual Symposium Proceedings. 2010:782–786. [PubMed: 21347085]
- Thomas SM, Mamlin B, Schadow G, McDonald C. A successful technique for removing names in pathology reports using an augmented search and replace method. Proceedings AMIA Symposium. 2002:777–781. [PubMed: 12463930]
- Van Horn JD, Ball CA. Domain-specific data sharing in neuroscience: what do we have to learn from each other? Neuroinformatics. 2008; 6(2):117–121. [PubMed: 18473189]
- Willison DJ, Emerson C, Szala-Meneok KV, Gibson E, Schwartz L, Weisbaum KM, Fournier F, Brazil K, Coughlin MD. Access to medical records for research purposes: varying perceptions across research ethics boards. Journal of Medical Ethics. 2008; 34(4):308–314. [PubMed: 18375687]
- Yue JK, Vassar MJ, Lingsma HF, Cooper SR, Okonkwo DO, Valadka AB, Gordon WA, Maas AI, Mukherjee P, Yuh EL, Puccio AM, Schnyer DM, Manley GT, Investigators T. R. A. C. K.-T. B. I. Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. Journal of Neurotrauma. 2013; 30(22):1831–1844. [PubMed: 23815563]
- Yuh EL, Mukherjee P, Lingsma HF, Yue JK, Ferguson AR, Gordon WA, Valadka AB, Schnyer DM, Okonkwo DO, Maas AI, Manley GT, Investigators T. R. A. C. K.-T. B. I. Magnetic resonance imaging improves 3-month outcome prediction in mild traumatic brain injury. Annals of Neurology. 2013; 73(2):224–235. [PubMed: 23224915]