Protecting Patient Identity in
SHRINE Without Lockout

David F. Walend, MS; Benjamin D. Carmen, BS; Marc Ciriello, BA; Isha Johri, BS; William Simons, MS; Bhanu Bahl, PhD; Ty P. Coghlan, BS Candidate
Harvard Clinical and Translational Science Center, Harvard Medical School, Boston, MA

Purpose

The Shared Health Research Information Network (SHRINE) helps researchers discover patient populations to support their research efforts. These populations are identified by querying networked i2b2 (Informatics for Integrating Biology & the Bedside) databases that contain data loaded from electronic medical records. By querying a SHRINE network, researchers are able to obtain aggregate counts of patients that match selected criteria.

To protect against patient re-identification, the SHRINE software injects noise into these aggregate counts and applies an automatic “lockout” of a researcher based on query behavior. Lockout prevents access to the network, but is not specific to researcher behavior inconsistent with network governance. In SHRINE version 1.19, we included a first version of a Data Steward Application (DSA) that allowed the “data steward” to approve or reject researcher-proposed query topics and review query behavior via a UI rather than looking at query logs.

Method

We examined SHRINE’s protection mechanisms and DSA in version 1.21 and identified gaps:

- Lockout was not specific, required coordination to reset: Researcher lockout produced false positives and unlocking a researcher involved at least three people and several days, costing network staff time and hindering network access.
- Researcher queries were cumbersome to review, easy to neglect: The DSA limited data stewards to reviewing one query at a time, and included no prompts for review based on usage of the network.
- Lockout did not prevent all bad behavior: SHRINE’s leaders were concerned about other malicious researcher behaviors, such as off-topic queries and data mining attacks.
- Researchers self-report lockout.

Result

We decided to shift the focus from disruptive, automatic protections like lockout while still providing protection against unintended network use. In SHRINE version 1.22, we added:

- New view of researcher behavior: A visualization of query term frequency by researcher and query topic. (example 1 and 2, below)
- Improved obfuscation: An increase in the default Gaussian noise injected from a sigma of 1.3 to 6.5, and query results rounded to the nearest 5 (configurable).
- Prompts for data steward review: An email reminder to examine each ACTIVE researcher’s queries every 30 queries, or at least every 30 days.

Conclusion

This multi-dimensional approach will address the concerns about re-identification and provide a way to for data stewards to detect off-topic queries without interfering with researchers using a SHRINE network.

Future Work

In future releases of the SHRINE software, we plan to improve the ease of use of the query view. Particularly, we hope to provide tools to support data stewards without a biomedical research background.