**ABSTRACT**

Distributed health data networks use distributed databases for efficient, privacy-protecting, and effective public health research and surveillance activities. Distributed regression analysis (DRA) is a novel analytic method that does not require transferring of patient-level data in multi-database studies but produces results statistically equivalent to those from pooled patient-level data analysis. The execution of DRA has been largely manual and labor-intensive. We describe a new approach to conduct automated DRA in the FDA’s Sentinel system, a distributed network using multiple electronic health data sources for medical product safety monitoring.

**OBJECTIVE**

- Implement a method within the existing PopMedNet™ (PMN) open-source platform used in Sentinel to allow automated, iterative, privacy-protecting, and scientifically accurate DRA in a real-world setting.

**BACKGROUND**

- Distributed databases enable efficient, privacy-protecting, and effective public health research and surveillance activities.
- DRA is a novel analytic method that does not require transferring of patient-level data.
- DRA produces results statistically equivalent to pooled patient-level data analysis.
- To date, DRA has been largely manual and labor-intensive.

**IMPLEMENTATION & RESULTS**

- PMN software development was an iterative process where the implementation ensured that the functionality developed within the PMN code base would not impact existing Sentinel workflows or system functions.
- We developed and validated PMN’s ability to perform regression analysis using only summary-level intermediate statistics and produce statistically equivalent regression parameters as pooled individual-level data analysis.
- Iterations driven by trigger files that indicate the start or end of a process, used by SAS and PMN.

**METHODS**

The project had two work streams:

- Work stream 1: Develop DRA analytic code in SAS for multivariable-adjusted regression analysis.
- Work stream 2: Enhance PMN to process DRA automated communication cycles within the distributed network.

**CONCLUSION**

- The DRA analytic code demonstrated reliable, accurate results when applied to test data sets; next steps include testing with Sentinel production data.
- Trigger files created and processed by the DRA analytic code and PopMedNet drive the automation and integration for successful DRA.
- This work can be leveraged in the future for DRA in Sentinel and other distributed health data networks.
- The functionality is agnostic to statistical software and can be extended to R and other software.

**DISCLOSURE STATEMENT**

- Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I.
- The authors have no relationships to disclose.