testing of novel software solutions. **Methods:** We convened a working group consisting of IT and research personnel with expertise in information security, database design/management, web design, EHR programming, and health services research. The working group was tasked with developing a research IT environment to accomplish two objectives: 1) maintain network/data security and regulatory compliance; 2) allow researchers working with external vendors to rapidly prototype and, in a clinical setting, test web-based tools. **Results:** Two parallel solutions, one focused on hardware, the second on oversight and management, were developed. First, we concluded that three separate, staged development environments were required to allow external vendor access for testing software and for transitioning software to be used in a clinic. In parallel, the extant oversight process for approving/managing access to internal/external personnel had to be altered to reflect the scope and scale of discrete research projects, as opposed to an enterprise-level approach to IT management. **Conclusions:** Innovation in health services software development requires a flexible, scalable IT environment adapted to the unique objectives of a HSR software development model. In our experience, implementing the hardware solution is less challenging than the cultural change required to implement such a model and the modifications to administrative and oversight processes to sustain an environment for rapid product development and testing.

C-D3-01:

Developing and Implementation of Secure Linear Regression on Distributed Databases

Fang Zhang, PhD, *Harvard Pilgrim Health Care*; Lingling Li, ScD, *Harvard Pilgrim Health Care*; Ken Kleinman, ScD; Jeff Brown, PhD, *Harvard Pilgrim Health Care*; Kimberly Lane, MPH, *Harvard Pilgrim Health Care*; Richard Platt, MD, MS, *Harvard Pilgrim Health Care*

Background/Aims: Improving access to multiple sources of electronic health information is crucial for population-based health care research. But it also raises serious issues relating to data confidentiality and security. We propose to develop methods and tools (e.g., SAS macros) to enable multivariate analyses to be conducted using multiple datasets that are stored separately without exposing any patient level information. Phase I of this project focuses on linear regression. **Methods:** A two-step process is carried out by SAS macros. The macros will be installed on both the central and site computers that hold the analytic datasets. After the central computer initiates the process, each site computer executes the SAS program locally and submits intermediate summarized statistical results to the central computer. The central computer then combines the intermediate results, and computes the parameter estimates. A second round of program execution delivers the variance/covariance estimates of the parameter estimates. **Results and Conclusions:** Simulated separated data sets were used to test the macros. The results were identical to the outputs from the SAS PROC REG procedure when the data sets were combined. The macro enables remote execution of multivariate linear regression analysis within the context of a multi-institutional distributed research network. The successful implementation of the secure statistical analysis within a distributed database model will overcome one of the largest hurdles in fully developing such a model, namely, the current need to combine data for multivariate analyses. Obviating this need will save the time and resources currently spent developing data-sharing agreements and will greatly minimize the patient confidentiality and proprietary concerns

C-D3-02:

Syndromic Surveillance Reporting Via CDC's Public Health Information Grid

Michael Klompas MD, MPH, *Harvard Pilgrim Health Care*; Ross Lazarus, MBBCh, MPH, *Harvard Pilgrim Health Care*; Gillian Haney, MPH, *Harvard Pilgrim Health Care*; Katherine Yih, PhD, *Harvard Pilgrim Health Care*; Richard Platt, MD, MSc, *Harvard Pilgrim Health Care*

Background: CDC is building a public health information grid to connect public health stakeholders nationwide. The grid will enable efficient and secure sharing of data and applications between public health authorities at the national, state, and local levels as well as researchers and other interested parties (JAMIA 2008;15:705). The grid may also support federated analysis of distributed data sources to detect clusters that might be invisible to smaller, isolated systems. Grid

protocols allow data owners to control data access. Broad sharing of data, however, requires a model to customize the level of patient identifiers seen by users depending upon their role. We describe here a work-in-progress implementation of an ambulatory syndromic surveillance module on the public health grid, including a model for sharing reports with different levels of granularity depending upon user privileges. **Methods:** We are installing a grid node at Atrius Health, a large multisite, multispecialty ambulatory medical practice serving over 600,000 patients predominantly in eastern Massachusetts. Standard CDC-defined ICD-9-and temperature-based syndrome algorithms developed for the National Bioterrorism Demonstration Project (MMWR 2004;53 Suppl:43) will be applied to ambulatory encounter data gathered nightly by the Electronic medical record Support for Public Health system (ESP, esphealth.org). ESP consists of a dedicated server populated with nightly extracts of comprehensive encounter data from the electronic medical record system of Atrius Health (JAMIA 2009;16:18). ESP currently analyzes this data to submit fully-identifiable case reports on patients with notifiable conditions to the state health department (MMWR 2008; 57:373). The syndromic surveillance addition to ESP will modify ESP's reporting module to vary the level of report granularity according to users' access privileges. ESP reports will include fully de-identified aggregate counts stratified by time period and zip code for basic grid users; semi-identified unit records with age, gender, and geo-code for intermediate users; and fully identifiable unit records with patient names and addresses for highly privileged users such as the state health department. **Results:** Installation is currently underway. **Conclusions:** Controlled distribution of syndromic surveillance data through the CDC's nascent grid will help develop the emerging paradigm for sharing public health information and applications.

C-D3-03:

Developing a Distributed Research Network to Conduct Population-based Studies and Safety Surveillance

John H Holmes, PhD, FACMI, *University of Pennsylvania School of Medicine for the HMO Research Network CERT DeCIDE Center Distributed Research Network project team*

Introduction: The overall objective of this project is to design a scalable distributed research network (DRN) that supports secure data analyses for research on the comparative effectiveness and safety of therapeutics. The plan for achieving this objective includes designing an information system architecture, implementation of a prototype of the specified system architecture, and completion of a proof of concept pilot project conducted in partnership with the Agency for Healthcare Research and Quality (AHRQ). Methods already implemented by the HMORN provide the foundation for the development of this DRN. These include established procedures for: 1) use of decentralized data files that remain in possession of the covered entities; 2) governance methods emphasizing local control and decision-making about participation in research projects; 3) transparency in choosing projects, protocol development, and communication of results; and 4) best practices for preserving privacy, ensuring data integrity, and accurate data interpretation. **DRN System Architecture:** The DRN architecture is being modeled as a federation of separate, heterogeneous databases that will be accessible through a common exchange model, implemented as a common query system (CQS). Under the common exchange model, heterogeneous data and systems do not need to be physically integrated, but the model affords interoperability, such that users can access these data resources using a single interface using a common syntax and semantic representation of data elements and coding schemes. This is a distinct advantage for the DRN, since there will be different data resource providers, some of which place strict constraints on the use of data. The common exchange model obviates the need for users to learn the syntactic and semantic characteristics of the resources available across the entire DRN. **Prototype Development:** The DRN is being developed using a rapid prototype life-cycle. This user centered approach employs incrementally-developed system prototypes such that all participants become stakeholders in the development process and have access to increasingly complex, concrete representations of the final system.