Information Technology
38

The Cancer Research Network's Virtual Data Warehouse

Gene Hart, Mark Hornbrook, Terry Field, Edward Wagner, and
members of the CRN Scientific Data Resources CoreSDRC group

Background: The Cancer Research Network's, a consortium of 11 HMO Research Network HMOs is in the process of building a (CRN) Virtual Data Warehouse (VDW) will to encompass eligibility, clinical, utilization, and socio-demographic data across all participating health plans the HMOs. We seek our goals are to reduce enable more efficient programming and data processing costs for multi-site projects, and to better build on the lessons learned from previous studies, and programs written, on previous projects., and to stimulate new proposals by reducing data extraction barriers.

Methods: The CRN Project Leaders Forum (PLF) sets research priorities. zes subject areas; the CRN Scientific Data Resources Committee (SDRC) develops a standard set of variable names and structures that can be used across the HMOs. The CRN Site Data Managers at each HMO writes programs to map their local data into the standardized set of names and structures. We have developed standard editing routines to detect problems in the data. The users of the VDW are expected to be any of the SAS programmers at the HMOs. The CRN Data Warehouse is "Virtual" because the data stay at the individual HMOs and. They are not centralized until a specific research use has received IRB/HIPAA approvals.

Results: There is considerable variability exists among HMOs in their implementingation of the VDW this concept. Some HMOs have disk space available and are making permanent copies of their data using the new structures. Others write programs, such as SAS Data Views, that make transient copies of the standard data structures only when required. Because of data volume problems some HMOs preselect their data to a population at risk before attempting to run a standard program.

Conclusions: SIt is simpler for the smaller HMOs to create and run programs using the VDW for their entire memberships and can turn around simple requests running programs on the VDW within a matter of hours. Very lthan it is for the larger HMOs. Since the larger HMOs have greater amounts of data they face substantial complexities in have more issues adopting the VDW modell. Some of the smaller HMO.s can turn around simple requests running programs on the VDW within a matter of hours.

While everyone can agree on the benefits of more efficient access to standardized datasets are clear, but a consistent set of concerns has been raised. Surprisingly, the ability to turning around a data requests very quickly presents some riskshas some process downsides. Previously in the past, thedata processing programming process required enoughtook enough time forthat local IRB issues to bewere settled beforeprior to completion of programming was finished.. With the ability to turn around requests in a day means that users there is a need to en put processes in place to make sure that IRB/ and HIPAA requirements are met before the local programmer runs the program and distributes data are distributed to other research centers.