

Calculating the Benefits of a Research Patient Data Repository

Ruth Nalichowski MBA, Diane Keogh, Henry C. Chueh MD, MS,
and Shawn N. Murphy, MD, Ph.D.

Massachusetts General Hospital and Partners Healthcare Inc. Boston, MA

The Research Patient Data Repository (RPDR) is a clinical data registry that gathers medical records from various hospital systems and stores them centrally in one data warehouse. Research investigators can obtain aggregate total of patients that meet specific query criteria and can obtain patient identifiers and complete electronic medical records through the RPDR with IRB approval. The existence of the RPDR is a critical resource to the Partners HealthCare System research community and supports many millions of dollars in clinical research. We have calculated the benefits of such a research system to help justify its establishment in other healthcare entities.

The purpose of the RPDR is to provide clinical investigators with the ability to perform population queries to support grant applications, for pre-trial patient cohort identification, and for hospital education and outcomes research. Investigators can easily query through millions of medical records, which are coded by patient diagnosis, medication, procedure, laboratory test, microbiology services, vital signs, molecular biology, demographic, and visit information. Through a web friendly visual query tool patients who meet the query criteria, as determined by the investigator, are returned as aggregate totals. With approval from the Institutional Review Board (IRB), researchers can obtain identified medical record information about the patients.

The RPDR was first released to a pilot set of users in December 1999, and was released in full production in February 2002. It has experienced a steady 20-30% increase of in the number of users performing queries and overall utilization each year. In 2005, there were 1073 users of the system, 308 of whom were new that year. The RPDR organizes the users into project teams both for the purpose of allowing queries to be shared, as well as controlling the release of patient identifiers (1). There were 245 active teams who in total received 578 sets of data consisting of the release of over 3.4 million patient records for research purposes.

The operational budget to maintain the RPDR was \$750 thousand in 2005. This does not include the current or initial development costs which vary by specific project approval but initially consisted of

approximately \$5 million over a 4 year period (during which no operational funding was available).

Due to the significant cost in operating and developing the RPDR, it was important to evaluate the systems' qualitative and quantitative benefits to the research community. The qualitative benefit of the RPDR is easier to identify than the quantitative benefit. Unfortunately in medical informatics, rarely do we know the benefit in Return on Investment with absolute certainty. Often a benefit is a combination of "fuzzy" ill defined concepts. (2).

The qualitative benefit of the RPDR focuses on the ability of an investigator to identify populations which are based on patients' medical records. Recruiting patients to participate in a research study by using the RPDR results in a higher enrollment rate versus traditional paper and radio advertising because a sample population of potential cohorts can focus recruitment based on the query criteria. The average cost for recruitment provided by sponsors, which includes advertising and remuneration, is approximately \$250 per subject. By adding up the subjects recruited by those that used the RPDR, assuming 50% of the approved numbers of subjects are enrolled in the study and 50% of those were recruited by the RPDR, approximately \$7 million is attributable to the use of the RPDR.

Utilization of the RPDR steadily increases each year and approximately 40% a study staff on all RPDR registered grants are RPDR users. In 2005, 40% of all grants registered with the RPDR obtained identified data. By taking into consideration user survey results and the number of requests for identified patient medical records per grant, the number of grants that are critically dependent on the RPDR range from 28% - 50%. Estimated money in 2005 alone that is funded by sponsors to grants that are critically dependent on the RPDR ranged from \$20.7 - \$30.7 million. The total funding of these critically dependent grants receiving data in 2005 ranged from \$94 - \$136 million.

1. Murphy, S.N., Chueh, H. A Security Architecture for Query Tools Used to Access Large Biomedical Databases. Proc AMIA Symp 2002, pages 552-556.
2. East, T.D. Return on Investment (ROI) Analysis in the Face of Uncertainty. Proc AMIA Symp. 2000, pages 234-9.