

Developing a future-proof database for the European Randomized Study of Screening for Prostate Cancer (ERSPC)

Tim Hulsen¹, Wim van der Linden¹, Cees de Jonge², Jonas Hugosson³, Anssi Auvinen⁴, Monique J. Roobol⁵, on behalf of the ERSPC consortium

¹Philips Research, department of Professional Health Solutions & Services, Eindhoven, the Netherlands

²Philips Research, department of Digital Platform Solutions, Eindhoven, the Netherlands

³University of Gothenburg, department of Urology, Gothenburg, Sweden

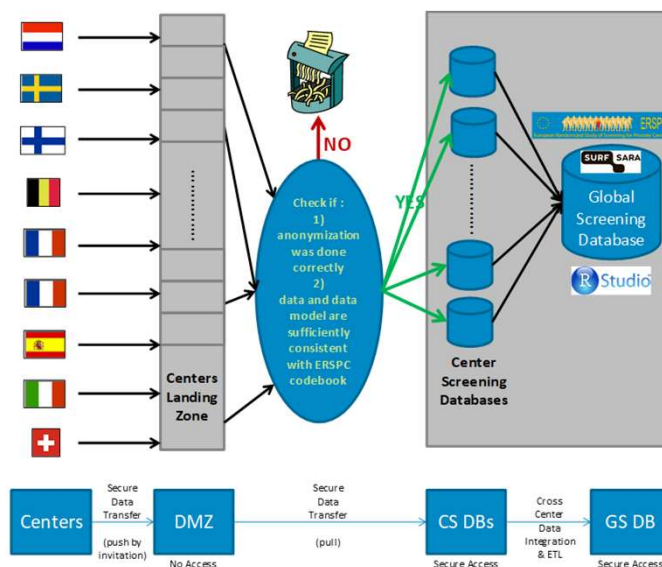
⁴University of Tampere, department of Epidemiology, Tampere, Finland

⁵Erasmus MC, department of Public Health, Rotterdam, the Netherlands



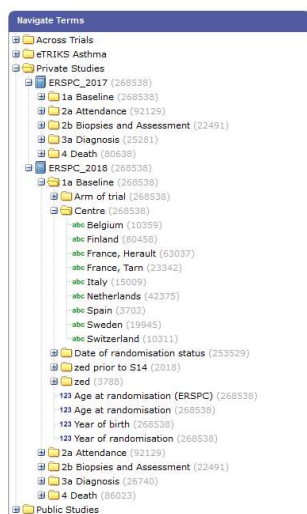
Introduction & Objectives

The European Randomized Study of Screening for Prostate Cancer (ERSPC)¹ is a combined and integrated study performed by 9 institutions in 8 European countries (the Netherlands, Sweden, Finland, Belgium, France, Spain, Italy and Switzerland), existing since 1992. It is the world's largest randomized prostate cancer screening study, and is designed to investigate whether early detection and treatment of prostate cancer might reduce disease-specific mortality and also help to identify men at risk. At its 25th anniversary, the global ERSPC database has been curated and re-organized. The database is accessible for analysis of the integrated datasets by all participating institutes.



Materials & Methods

The global ERSPC database IT infrastructure is based on the tranSMART prostate cancer instance of the TraIT IT infrastructure developed by Philips within the CTMM-TraIT, CTMM-PCMM and Movember GAP3 projects. The tranSMART tm-curation pipeline has been used to curate the full dataset, and data has been loaded into the tranSMART platform using the tranSMART-batch pipeline. This infrastructure also offers support for collecting and combining yearly updates of the various large, longitudinal datasets from the participating institutes. Transfer of data takes place using the Secure Data Transfer tool provided by Philips.



Results

The global ERSPC database contains datasets of 268,539 patients, from nine institutes from eight countries. The clinical data has been gathered using a common data model, specifically designed to answer the research questions defined by the principal investigators at the start of the project. Data is organized in five main categories: Baseline, Attendance, Biopsies & Assessment, Diagnosis and Death. Attendance and Biopsies & Assessment each contain data from multiple rounds and visits. This data can be browsed through the web-based tranSMART platform, which is only accessible for a selected group of users. tranSMART supports a number of statistical analyses, such as survival analysis, logistic regression and correlation analysis. The ERSPC tranSMART instance is connected to R-studio and STATA to enable the statisticians to execute their own scripts on the database.

Conclusions

We show here that it is feasible to curate a global screening database that has been existing for 25 years, and to make it available in a future-proof statistical analysis environment, which enables researchers to browse and analyse the clinical data in a secure and efficient manner. The IT infrastructure, that includes tranSMART, R-studio and STATA, will be used in future projects that handle large clinical datasets.