



CLINICAL TRIAL DATA CAPTURE IN TOLVEN

Introduction: Data Comparability from Bedside to Bench

Bridging the gap between clinical research and patient care is a challenge. Although clinical investigators and clinical care providers may be interacting with one another on a frequent basis and may be concerned with identical patient populations, clinical research applications have been remarkably separate from clinical care applications. This separation has generated needless extra work for all stakeholders and has contributed to the potential for mismatch between clinical care and clinical research data.

It is important to understand the difference between data use, such as for a given study, and data collection, such as by a provider of care or directly by the patient. This detail has a dramatic impact on the reusability and durability of instances of metadata definitions. As a simple example, a provider or patient typically collects the patient's date of birth. Age can be easily determined at any time based on date of birth. Further, each different study can group patients by age range using patient age computed from date of birth without having to define specific age ranges. This is critical to semantic interoperability, since exchanging age or age ranges is not likely to yield comparable results except in very narrow contexts.

TRANSCEND and I-SPY

The I-SPY trial (<http://tr.nci.nih.gov/iSpy>) is a multi-center study which combines biomarkers to predict response to therapy in patients with locally advanced disease undergoing neo-adjuvant treatment. I-SPY uses an adaptive design where treatment response is used to randomize future subjects, thus dramatically accelerating drug evaluation in breast cancer. Information learned in I-SPY will lead to tailored and accelerated evaluation of therapies based on genomic signature.

As it includes the integration and analysis of diverse data types, I-SPY involves a number of informatics challenges, including MRI, gene expression, immuno-histochemistry, fluorescent in-situ hybridization (FISH), and other emerging proteomic tissue signatures. Furthermore, the multi-center nature of the study and the need for highly specialized testing requires a robust distributed system capable of tracking patient specimens and aggregating diverse data types in an integrated visual representation.

TRANSCEND (**TRAN**slational Informatics **S**ystem to **C**oordinate **E**merging Biomarkers, **N**ovel Agents, and **C**linical **D**ata) addresses the next phase of the I-SPY informatics effort, and has focused primarily on the integration of a Tolven electronic clinical record system with the I-SPY research data infrastructure. An ongoing key objective is to demonstrate the use of standards to integrate the Tolven clinical record system within the I-SPY infrastructure and to utilize clinically-driven data collection in support of translational research.

The TRANSCEND project uses Tolven software to document and store the information needed by both care providers and researchers for patients with advanced breast cancer, and therefore erases the historical separation between clinical care and clinical research informatics activity ([https://cabig.nci.nih.gov/.../cabig.../1-Hogarth-2009 caBIG 2 final.ppt](https://cabig.nci.nih.gov/.../cabig.../1-Hogarth-2009_caBIG_2_final.ppt).)

TRANSCEND involves the exchange of de-identified clinical information across a consortium of academic medical centers and research laboratories. TRANSCEND therefore addresses such important issues as patient privacy, interoperability, and clinical data standards as well as the transformation of clinical care information into clinical research data. The TRANSCEND project is housed at the University of California San Francisco's Helen Diller Family Cancer Center and is funded by the National Cancer Institute.

The TRANSCEND project provides a blueprint for bridging the gap between clinical research information, and clinical care information. Clinicians from the University of California San Francisco (UCSF) and the University of Pennsylvania have supplied the major portion of the clinical domain input. The software components of the project are being developed by Tolven and are being continually contributed to open source.

Tolven's Role

The Tolven team has built a federated system for data collection, adaptive randomization, and biospecimen tracking within the I-SPY 2 trial. The system uses Tolven components and prevailing health data standards, thus enabling wide scale adoption of this infrastructure for future cancer trials. The crux of the system is to enable the transfer of clinical information to standard formats for clinical trial registration and randomization, and to enable the integration (use and testing) of biomarkers to tailor care.

Engineers have developed electronic case report forms (CRF) using the Tolven web-enabled clinical information system infrastructure. This involves two modes of CRF creation: (1) web-based CRF forms and (2) assembly of completed CRFs from data entered by clinical staff into a Tolven clinical record application. I-SPY 2 participating institutions have a choice of either mode. At least two institutions, UCSF and the University of Pennsylvania, are committed to the auto-assembly mode, which involves clinical staff entering patient data into the Tolven clinical information system in their respective clinical areas.

A central feature of Tolven is its rule engine and the way it works: All data destined for a patient's medical record (regardless of source) are processed by rules. Simple rules will, for example, put issues identified as problems onto the problem list. Or the most recent glucose reading for a patient might be added to a diabetes registry.

Rules can also route certain data to other "accounts," such as drug orders to the pharmacy, lab orders to the lab, etc. Some of these rules might route information conditionally. For example, if the patient has consented to participate in a particular study, then selected data (identifying the patient or not) may be routed to that study.

Rules allow each study to address very different questions to a normalized medical record. For recruitment, many different studies can inspect the same set of medical data to determine if the patient qualifies or not for that study. The rules can even encourage the elicitation of data that is missing or out of date. From simple answers to simple questions in the clinical record, Tolven can produce complex unambiguous answers to complex questions. During any study or clinical trial, patient data does not need to be entered twice - once for the medical record and again for the study – thus avoiding the inevitable transcription errors.

In addition, the clinical data base will allow research personnel the opportunity to ask different questions that had not been anticipated in the initial study design. For instance, if it became important to re-examine research eligibility categories and adjust the age upward or downward, the clinical data base would permit those altered queries without the need to re-enter data, since the rules could be adjusted to reflect the new queries.

The Tolven solution includes a rules engine that utilizes the common rules language, OPS5. OPS5-based rules have been developed to enable the auto-assembly CRF construction from clinical data. The rules are necessary in order to have the system help establish the suitability and completeness of captured clinical data as appropriate “answers” for the questions in the case report forms.

The I-SPY and TRANSCEND project has also required the establishment of a core set of common data elements for breast cancer. Analysts have derived needed elements from accepted standard terminologies (chiefly SNOMED) and have mapped them to the NCI caDSR (National Cancer Institute Cancer Data Standards Registry and Repository). Any needed elements that were not in the caDSR are continually submitted for proposed inclusion in the caDSR.