Supplemental Figure 1: Cost Difference Calculation

Supplemental Figure 2: Decision Tree—An abbreviated decision tree graphically summarizes the model used to compare data acquisition, cleaning, and linkage under a standard trial design versus a registry-based trial with data accessed at each individual center via a study-specific query.

Supplemental Figure 3: Summary of Semi-Structured Interview—Open-ended questions were asked to the following groups of people on the following subjects.

Supplemental Figure 1: Cost Difference Calculation

Cost difference, Δ_C, were defined as follows:

$\Delta_C = C_{Reg} - C_{Std}$

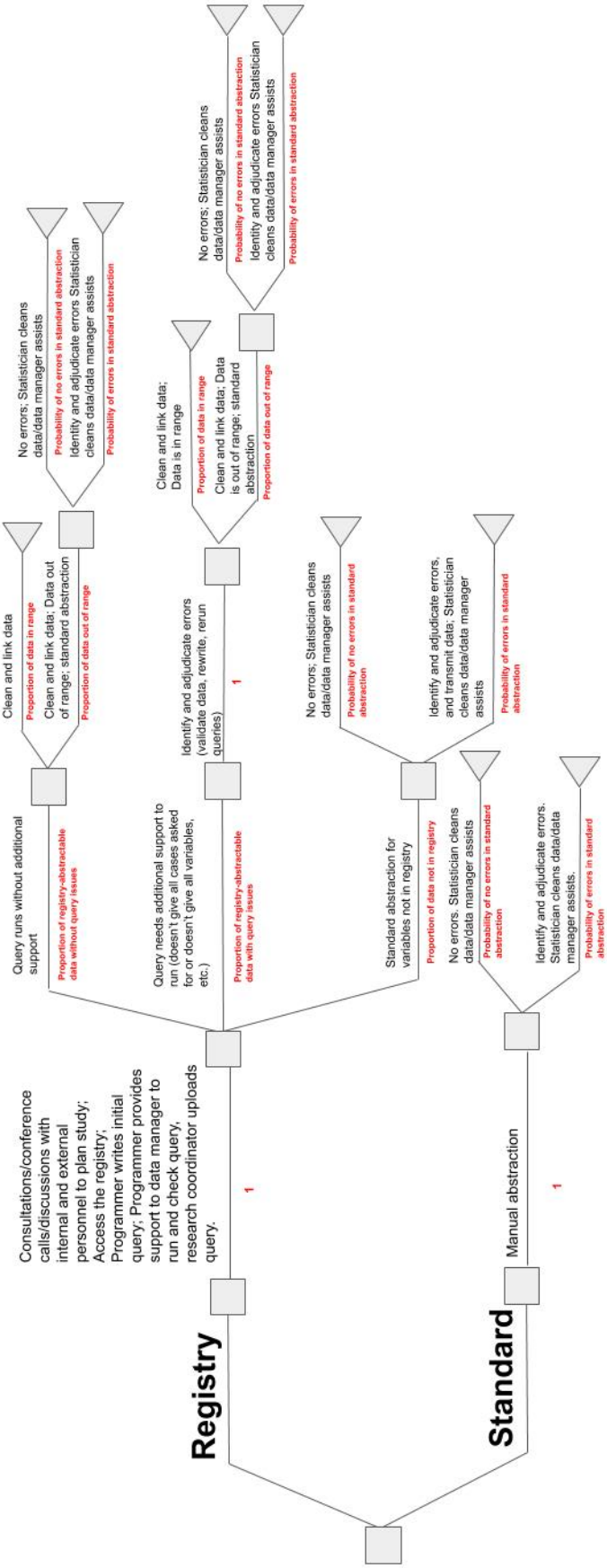
Where C_{Reg} = Costs of the registry-based trial,
and C_{Std} = Costs of running the same trial with manual data collection.

$C_{Reg} = C_{Shared} + C_{FixedReg} + C_{VarReg} + C_{AbstReg}$

$C_{Std} = C_{Shared} + C_{AbstStd}$

Where $\mathbf{C_{Shared}}$ = All costs common to both trial types. These costs cancel when C_{Std} is subtracted from C_{Reg},
$\mathbf{C_{FixedReg}}$ = Fixed costs associated with accessing registry data (range includes zero) and cleaning and linking these data,
$\mathbf{C_{VarReg}}$ = Variable costs associated with planning, writing, running, checking, and uploading initial data query,
$\mathbf{C_{AbstReg}}$ = Variable costs associated with the manual abstraction, uploading, cleaning, editing, and linking of data not accessible through the registry,
$\mathbf{C_{AbstrStd}}$ = Variable costs associated with manual abstraction, uploading, cleaning, editing, and linking of all data in the standard clinical trial.

Supplemental Figure 2: Decision Tree



Supplemental Figure 3: Summary of Semi-Structured Interview

Research Coordinators

- The average time it takes to abstract one variable per patient
- The average time it takes to upload a query
- The percentage of patient charts that had to be revisited in order to adjudicate out of range input errors

Data Managers

- The average time it takes to run a query provided by a programmer in the registry case
- The average time it takes to run and check a query in the standard case
- The average time it takes to clean data in the standard case

Statisticians

- The average time it takes to clean data in the standard case
- The average time it takes to clean and link data in the registry case
- The likelihood that the queried data will be out of range and need to be abstracted again manually

Programmers

- The average time to write the initial query
- The average time to implement the query per center
- The average support time per center
- The likelihood that any one center will need additional query support or a query rewrite
- The average time to edit the query per center