# Power Analysis

Statistical inferences from our proposed research will be based on model-based estimates of risk factor effects on ACLR outcomes. Therefore, the objective of the power analysis was to calculate the precision with which our models could estimate risk factor parameters, given a realistic study sample size. Thus, we used a simulation-based approach to calculate the power of detecting moderate effects for both fear of movement/reinjury (as measured by TSK) or self-efficacy (as measured by GSES) on the appropriate outcome variables corresponding to each of the stated study aims. For each model, we generated 1000 simulated datasets and used a simplified version of the models described below to generate estimates of risk factor parameters for each dataset. Power was estimated by calculating the proportion of time the corresponding 95% Bayesian credible intervals (BCI) excluded zero, for both fear of movement/reinjury and self-efficacy.

Datasets were simulated using parameter values either estimated from existing data or derived from extant literature. Covariate parameter estimates, covariances and standard errors were estimated from linear mixed-effects models using a sample of 155 patients in the consortium database. Parameters for risk factor effects were set to plausible values (correlation coefficients, Table 1) based on published studies relating either TSK or GSES to each of the ACLR recovery measures.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Aim 1** |  | **Aim 2** |  | **Aim 3** |  |
| **Risk Factor** | IKDC | KOOS QOL | Single hop | Crossover hop | KOOS spts/rec | Marx |
| TSK | -0.22 | -0.22 | -0.3 | -0.3 | -0.45 | -0.45 |
| GSES | 0.25 | 0.25 | 0.4 | 0.4 | 0.6 | 0.6 |

***Table 1****: Correlation coefficients for risk factor effects*

Correlation coefficients ( were converted to linear model parameters ( via:

where and are the standard deviations of the ACLR recovery response and risk factor, respectively. These values were combined with the estimated covariate parameters in a single generative model to produce simulated data. Each dataset consisted of 380 samples, which corresponds to a realistic estimate of our study sample size. In addition, we randomly allocated an average of 15% of the sample to be lost to followup in each simulation.

We fit Bayesian multivariate linear models to each dataset, calculating posterior 95% credible intervals for both risk factor effects in each case. Each model was fit using Markov chain Monte Carlo methods, with diffuse normal priors (mean=0, variance=1000) for each regression parameter. Models were run for 10,000 iterations, with the first 9000 samples discarded as burn-in; models were automatically monitored for convergence using the Gelman-Rubin statistic (Gelman & Rubin, 1992).

Calculated power was extremely high, with all but two of the 12 risk factors yielding 100% power under this simulation scenario. Only the risk of movement/reinjury effects on single hop and crossover hop scores had lower power, with estimates of 66% and 53%, respectively.

References

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science. A Review Journal of the Institute of Mathematical Statistics*, 457–472.