Title: Inter-rater agreement of the Central Adjudication Committee of the PARITY trial in the diagnosis of surgical site infection following limb-salvage surgery with endoprosthetic reconstruction

Authors: James Nuttall, Nathan Evaniew, Patrick Thornley, Anthony Griffin, Benjamin Deheshi, Tim O'Shea, Lor Randall, Robert Turcotte, Peter Ferguson, Jay Wunder, Mohit Bhandari, Michelle Ghert

Background: Endoprosthetic limb reconstruction for bone tumors is associated with a high risk for surgical site infection and ultimate amputation. The diagnosis of surgical site infection is challenging, and experts agree that there is no gold standard test. In large clinical trials, blinded Central Adjudication Committees (CACs) are often utilized to minimize the variability associated with having multiple investigators at multiple sites and to minimize bias in outcomes assessment. The authors are currently involved in the large multicenter international Prophylactic Antibiotic Regimens in Tumor Surgery (PARITY) trial. The aim of this study was to examine the level of agreement of the PARITY CAC in diagnosing surgical site infection.

Methods: The CAC for the PARITY trial is a five-member panel of orthopaedic surgeons and an infectious disease specialist. The PARITY CAC members were asked to adjudicate 29 non-PARITY cases of lower extremity endoprosthetic reconstruction. Cases were identified from the databases of two separate sarcoma centers, and short clinical vignettes including available recorded information were presented in a survey format. The panel members were asked to decide independently whether or not an infection was present and to classify any infection based on the Centers for Disease Control (CDC) criteria of surgical site infection (superficial, deep and organ space). In the case of insufficient information, panelists were able to indicate that they were 'unable to assess'. The level of consensus was recorded and Fleiss Kappa values were calculated to determine the reliability of the panel in identifying the presence of infection and the CDC type of infection. Combinatoric analysis of Fleiss Kappa value was used to calculate the smallest CAC panel size required to maximize reliability. The survey was repeated 3 months after initial responses had been collected. These data were used to determine intra-rater reliability using Cohen's Kappa. Finally, a meeting was held to establish a consensus opinion on each case that did not produce a full consensus between individual panelists. Discussions were analyzed and potential issues to be resolved for the PARITY trial were identified.

Results: All five panelists (full consensus) or four of five panelists (near consensus) agreed on the presence or absence of surgical site infection in 20 of the 29 cases. The Fleiss Kappa value was calculated as 0.44 (95% CI 0.35- 0.53), which is consistent with moderate agreement. When asked to classify each clinical vignette according to the CDC criteria of deep or organ space surgical site infection, panelists reached a full consensus in 12 of 29 cases and near consensus in 5 of 29 cases. This corresponded to a decrease in kappa agreement to fair, 0.35 (95% CI 0.28-0.42). When all possible responses were examined, the greatest statistical agreement was observed in the outcome of no infection, 0.61 (95% CI 0.49-0.72), consistent with substantial agreement, and any infection, 0.58 (95% CI 0.47-0.70) indicating moderate agreement. Following the second round of adjudication, intra-rater agreement was substantial (0.67, 95% CI 0.87-0.47) for presence or absence of infection and moderate for CDC criteria (0.57, 95% CI 0.75-0.39). Analysis of CAC size showed a stable maximum Fleiss Kappa of 0.46 (95% CI 0.50-0.35) at CAC sizes of 3 members or more. Sixty-two percent of all cases required consensus discussion. The average discussion time to reach a consensus was 3 minutes. The consensus of the panel differed from the majority individual opinion in 17% of all cases. In this minority of cases, the most common conclusion was that the CAC was unable to assess the presence or absence of infection or reach a categorization based on CDC criteria due to inadequate clinical information.

Conclusion: The PARITY CAC can agree on the presence or absence of surgical site infection following lower extremity endoprosthetic reconstruction, but that agreement on the level of infection is more challenging. The PARITY trial is powered for 'infection' vs. 'no infection', and the current study supports this statistical plan for the PARITY trial. These results support a minimum CAC size of 3 members. We identified a decrease in reliability when retrospective clinical data is used. The standardized data collection and data quality maintenance used in the PARITY trial will improve the reliability of the CAC.