PKIDS Trial Supplemental Analysis plans

Methods to control confounders. Covariates across treatments will be balanced using an inverse probability treatment weighting (IPTW) approach for multi-level data to evaluate overall treatment effect and heterogeneity of treatment effect for all outcomes.(52) Propensity score models will use random effects to account for correlation within surgeon- and hospital-levels. We will examine overlap of confounders across treatment groups by comparing estimated propensity scores, and examine the balance of confounders using standardized difference in mean before and after weighting.(53-55) Additionally, sensitivity analyses using both simulation-based approaches(56) and propensity score-based sensitivity functions(57) will assess the effect of potential unmeasured confounding on the estimates of the parameters of interest.

Process Evaluation. Surgery is a complex intervention influenced by patient and disease characteristics; surgeon experience and preferences; hospital resources; operative equipment; and procedural complexity, among other factors. A highly regimented surgical approach would neither reflect actual clinical care nor allow identification of operative techniques that could be modified to improve outcomes. Thus, a process evaluation will be incorporated, ensuring certain standard of care practices are being followed and recording key technical factors for each procedure. To this end, we will use a detailed data collection form that records standard operative processes and key steps for ureteroscopy, shockwave lithotripsy, and percutaneous nephrolithotomy. Ureteral injury following ureteroscopy will be assessed as described by Traxer(58), while the Guy's Stone Score will be used to assess stone complexity associated with percutaneous nephrolithotomy.(59) Additionally, we will record detailed information about the surgical indications, clinical urgency, equipment used, and the key surgical steps of the

procedure. A focused training program and review with PKIDS surgeons will ensure standardized reporting. Finally, surgeons will document the anticipated surgical plan and any variations thereafter (i.e. an unplanned staged ureteroscopy) will be recorded. Supplemental Table displays the varying surgical processes that are evaluated with each modality.

Supplemental Table: Process evaluation for ureteroscopy, shockwave lithotripsy, and percutaneous nephrolithotomy.

Surgical Process	Surgical Modality		
	Ureteroscopy	Shockwave Lithotripsy	Percutaneous Nephrolithotomy
Indication for Operation	X	X	X
Plan for Staged Intervention	X	X	X
Variation from Initial Surgical Plan	X	X	X
Surgical Equipment Used	X	X	X
Antibiotic Administration	X	X	X
Anesthetic Type[1]		X	
Use and Type of Intra-operative Imaging	X	X	X
Ureteral Access Strategies	X		
Shockwave Parameters		X	
Percutaneous Access Strategies			X
Safety wire utilization	X		X
Percutaneous Endoscopic Visualization[2]			X
Laser Parameters	X		X
Intra-operative Bladder Drainage	X	X	X
Post-operative Urinary Drainage[3]	X	X	X
Intra-operative Complications	X	X	X
1: Ureteroscopy and percutaneous nephrolithotomy are universally performed under general anesthesia at all sites 2: Including flexible nephroscopy and antegrade ureteroscopy 3: Including both bladder drainage, ureteral stenting, and nephrostomy drainage as indicated			

Preventing, monitoring, and handling missing data. To decrease missing data, we will implement the following measures: (1) monetary incentives for participants to complete PROs; (2) structured data in operative reports; (3) automated reminders in REDCap to prompt PRO completion; (4) coordinator-initiated contact if PROs are incomplete; and (5) text message communication, which we have found to be an effective means of communicating with adolescents with kidney stones. We will report the reason for dropout; who decided that the participant would drop out; and whether the dropout involved some or all types of participation. We will describe missing data, including a comparison of baseline characteristics of patients and PKIDS sites with and without missing data. Our data analysis will assume data missing at random, but we will also include predictors of dropout to relax assumptions on missing mechanism. For missing values in >5% of the outcomes, we will implement multiple imputation methods. We will perform sensitivity analyses to investigate possible informative dropout (for reasons related to the type of surgery) and examine the sensitivity of our assumptions about the mechanism of missing data.

Sensitivity analyses. Several sensitivity analyses will evaluate the robustness of model assumptions as follows: 1) changing outcome cutoff points (*e.g.* size of residual fragments); 2) changing the analytical method (*e.g.* different matching approaches rather than IPTW); 3) assessing the effect of potential unmeasured confounding on the estimates of the parameters of interest; 4) changing missing data handling methodology; 5) influence of minor protocol deviations; 6) management of outliers; and 7) impact of hospital site on outcomes. We will report the results in accordance with STROBE standards.(60)

Additional PRO Analysis. We will conduct longitudinal analysis of patient experiences assessed by PROs and fit WGLMMs to compare the slopes of the PROs over time. In order to explore the difference in the post-operative experience between patients who undergo a staged surgery and a single surgery, we will compare the PROs measured at 1 week after each of the two procedures in the staged surgery to the PROs measured at 1 week after the single surgery. Specifically, we will fit an extended weighted linear mixed effects models to test the difference in T-scores for each PROMIS domain. The extended models will account for the correlation between the two measures of PROs in patients with the staged surgery. Then, we will compare the trajectory of outcome from 1 week to 3 months after surgery between patients with the staged surgery and with the single surgery, using the generalized additive modeling. Random effects will be used to account for potential correlation among patients within the same hospital and surgeon. We will also control for potential patient-, hospital- and surgeon-level confounder effects in the regression models. Similar to the primary analysis, we will examine for heterogeneity of treatment effect of PROs by testing for the interaction terms between procedure and patient age and procedure and sex to identify associations that differ for subgroups. We will also identify modifiable surgical techniques that are associated with patients' experiences for each procedure (ureteral access sheath for ureteroscopy, shock frequency for shockwave lithotripsy, and size of access tract for percutaneous nephrolithotomy).

DVSS will be analyzed by comparing change in score from pre-treatment to post-treatment with comparisons across modalities, acknowledging a 2 point difference in DVSS score to be clinically significant. While QUIKSS will not be formally validated within the primary SAP, this questionnaire will be analyzed in several stages. First, we will assess the relationship between various PROMIS PRO domains and QUIKSS measured urinary symptoms. Second, we will

compare responsiveness of QUIKSS and DVSS to intervention. Third, we will explore responsiveness and correlation of QUIKSS with PROMIS and DVSS over time as a component analysis.