Urodynamic Standardization in a Large-Scale, Multicenter Clinical Trial Examining the Effects of Daily Tadalafil in Men With Lower Urinary Tract Symptoms With or Without Benign Prostatic Obstruction

Stephen R. Kraus,* Roger Dmochowski, Michael E. Albo, Lei Xu, and Claus G. Roehrborn

Aims: To present the methodology, standardization techniques, and results from post hoc test–retest reproducibility analyses for a large, placebo-controlled, multicenter trial, employing urodynamic studies (UDS) to assess the impact of daily tadalafil on men with lower urinary tract symptoms (LUTS) with or without benign prostatic obstruction (BPO). Methods: UDS implemented International Continence Society (ICS) Good Urodynamic Practice guidelines and standardized urodynamic and LUTS terminology. Further standardization procedures included: equipment calibration; a detailed procedure manual and centralized training; and implementation of a central reader. Measures included: monitoring of invalid studies, comparison of actual versus expected standard deviation (SD) for primary outcome (detrusor pressure at maximum urinary flow rate [p_detQmax]), and test–retest reproducibility of the placebo arm at baseline and endpoint. Results: Two hundred men with moderate to severe LUTS (baseline IPSS >13) at 20 sites were randomized to receive either tadalafil 20 mg or placebo. All men underwent non-invasive uroflow and pressure-flow studies. Numbers of invalid studies at baseline and endpoint were 9.3% and 0.6%, respectively. Variability of p_detQmax was lower than anticipated based on actual versus expected SD of 15 and 30, respectively. Correlation coefficients were very good for pressure-flow parameters including p_detQmax (r = .83). Conclusions: Multicenter clinical trials using urodynamic outcomes require additional standardized procedures to limit inter-site variability. By implementing centralized training with a detailed procedure manual and use of a central reader, we were able to limit common difficulties arising in multicenter clinical trials, as well as demonstrate good test–retest reproducibility of pressure flow measures. Neurourol. Urodynam. 29:741–747, 2010. © 2010 Wiley-Liss, Inc.

Key words: benign prostatic obstruction; lower urinary tract symptoms; multicenter trials; standardization; tadalafil; urodynamic study

INTRODUCTION

Despite the use of standardized urodynamic practices outlined by the International Continence Society (ICS), variability in how urodynamic studies (UDS) are performed still exists. Schafer et al.2 reported that UDS performed as part of the multicenter ICS-BPH Study were often difficult to interpret due to artifact, variability in the order or scaling of the traces, improper zeroing, and small scale print-outs. These inconsistencies, combined with the fact that individual interpretation of the results allows for latitude, can result in poor data reproducibility potentially limiting the ability to use standardized UDS in multicenter clinical trials. In an attempt to minimize these limitations, Lewis and Abrams3 described a standardized UDS protocol and advocated the use of central evaluation. However, in the current trial, we identified areas of variability that could still impact urodynamic observations, including a lack of standardized verbiage across sites, as well as inconsistencies in the training of urodynamic testers and interpretation of urodynamic data which could lead to low inter-rater reliability between local investigators and a central reader. In order to surmount these problems and minimize variability between sites, we developed standardized procedures in addition to the implementation of the ICS Good Urodynamics Practice Guidelines and Standard Terminology for the conduct and assessment of non-invasive uroflow (NIF) and pressure flow studies (PPS).

Recently, Hashim et al.4 reported a post hoc analysis of 35 placebo-treated patients in a multicenter clinical trial which demonstrated good test–retest reproducibility in urodynamic measures over 6 months. To assess the issue of data consistency, we applied a similar analysis to the current trial to examine the test–retest reproducibility of the UDS in this larger clinical trial over 12 weeks. The aim of this article is to present the methodology used in our attempt to reduce variability among sites, and provide the results of these post hoc analyses.

Conflicts of interest: Drs Kraus, Dmochowski, Albo, and Roehrborn were paid consultants and served on the advisory board for this study. Dr Xu and Mrs. Klise are employees of the sponsor. Grant sponsor: Eli Lilly and Company.

*Correspondence to: Stephen R. Kraus, MD, Department of Urology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, Mail Code 7845, San Antonio, TX 78229-3900. E-mail: krauss@uthscsa.edu

Received 27 March 2009, Accepted 18 August 2009

Published online 11 February 2010 in Wiley InterScience (www.interscience.wiley.com)

DOI 10.1002/nau.20825

© 2010 Wiley-Liss, Inc.
MATERIALS AND METHODS

The objective of this analysis is to assess the ability to obtain consistent data from standardized urodynamics in a multicenter clinical trial on men with lower urinary tract symptoms (LUTS) who may or may not have benign prostatic obstruction (BPO). Outcome parameters include: assessing the number of invalid studies, comparing expected versus actual standard deviation (SD) of the urodynamic primary outcome and examining the test–retest reproducibility of select PFS and NIF parameters.

This 12-week, multicenter clinical trial was designed to determine the effect of daily tadalafil 20 mg versus placebo on urodynamic parameters in men with LUTS. Subjects were randomized at 20 sites in the US and Canada. The trial’s primary aim was to compare the difference of change from baseline to endpoint in detrusor pressure at maximum urinary flow rate ($P_{\text{detmax}}$) between treatment groups. Secondary objectives included changes in additional PFS and NIF parameters, as well as symptom improvement as measured by the International Prostate Symptom Score (IPSS), which allowed for correlation of objective urodynamic findings with subjective symptom improvement.

This trial was sponsored and monitored by Eli Lilly and was designed in collaboration with a panel of urologists with expertise in the evaluation and treatment of men with LUTS. The protocol was reviewed by all applicable ethical review boards. The trial was conducted in accordance with the principles of the Declaration of Helsinki and Good Clinical Practices (GCP) and followed all applicable laws and regulations. Informed consent was provided by each subject.

Standardized Urodynamic Procedure

Site selection was based on adequate past experience with UDS in men with LUTS, provision of acceptable examples of urodynamic tracings, and completion of certified urodynamic training with subsequent certification. To minimize variability among sites, UDS methods were standardized in accordance with ICS guidelines, and all sites were required to adhere to a list of standardized annotations and verbiage (Table I).

Centralized Training

All urodynamic testers submitted sample tracings on men with LUTS which were reviewed by the central reader and trainer. If these tracings were deemed acceptable, testers were then required to attend centralized training which included lectures designed to review the protocol and potential UDS artifacts, as well as participate in UDS hands on training. Additionally, all testers were required to review a DVD of the urodynamic procedure being performed in accordance with the Urodynamic Procedures Manual. All sites received a copy of the DVD and manual for reference. Testers were required to submit two sets of NIF and PFS tracings of men with LUTS performed in accordance with the procedures manual which met minimum requirements, as verified by urodynamic physician experts. Only urodynamic testers who could demonstrate compliance with the UDS standardized procedures were certified.

Non-Invasive Uroflowmetry Procedure

Subjects were instructed to arrive with a full bladder. A NIF was conducted to assess maximum flow rate ($Q_{\text{max}}$), average flow rate ($Q_{\text{ave}}$), and voided volume ($V_{\text{comp}}$). If the NIF volume was <125 ml, it was considered invalid and a repeat was attempted. The subject voided in his normal position (standing or sitting). The channel order for signal configurations was standardized and an example of NIF in proper orientation is seen in Figure 1. Tracings were reviewed by investigators and testers at the end of each assessment to ensure minimum requirements were met. An invalid NIF assessment was defined as a voided volume <125 ml or an illegible tracing. This was followed by urethral catheterization to measure post-void residual volume and for performance of urinalysis. If no evidence of infection was present, the remainder of the UDS continued.

Cystometrogram and Pressure-Flow Procedure

All PFS were performed with fluid based manometers utilizing urethral and rectal catheters for simultaneous measurements of vesical and abdominal pressures. A 7 Fr dual-lumen urethral catheter and a standard commercially available closed-system rectal balloon catheter (to be instilled with 2–3 ml H2O) with externally based fluid filled transducers were required. No other types of catheters, including electronic or light sensing catheters, were allowed. Urodynamic testers were instructed to continuously record intravesical pressure ($p_{\text{ves}}$), intra-abdominal pressure ($p_{\text{abd}}$), and subtracted detrusor pressure ($p_{\text{det}}$) on computerized, multi-channel, urodynamic equipment throughout the filling and voiding phases. A medium infusion rate of 50 ml/min was required. The infusion rate could be decreased at the discretion of the tester if detrusor overactivity (DO) or other storage symptoms developed; if DO persisted, testers were instructed to stop filling until DO subsided and restart filling at a lower infusion rate. The channel order for signal configuration was standardized (Fig. 2A). Urodynamic testers were required to adhere to a list of standardized annotations and verbiage.

Urinalysis

Fig. 1. Sample of a non-invasive flow tracing, including orientation and scaling.
and endpoint NIF and PFS. Each tracing was annotated to record the subject's voiding position in order to confirm the same position was maintained. Prior to the initiation of the cystometrogram (CMG), all catheters were assessed to ensure proper functioning. This was accomplished by ensuring appropriate concordance between $p_{ves}$ and $p_{abd}$ with each cough, which was annotated at the peak of the pressure spike allowing the urodynamic tester to look for the dynamic response of the pressure channels. If $\geq 70\%$ concordance$^{5}$ was not achieved, the tester was trained to stop filling and troubleshoot the system. A CMG baseline pressure was obtained once the UDS was started and a discrepancy of no greater than $\pm 5$ cmH$_2$O was tolerated for the baseline detrusor pressure. Beginning at a bladder volume of 100 ml and continuing throughout the filling, proper placement and function of the catheters and transducers was verified with coughs at intervals of 50–100 ml. All coughs were monitored to ensure appropriate concordance as described above, with troubleshooting performed as necessary. Urodynamic testers were trained to watch for evidence of DO. For the purposes of this clinical trial, in order to standardize and improve the likelihood of detection, DO was annotated for detrusor

Fig. 2. A: Sample of cystometrogram/pressure-flow study tracing, including orientation and set up. The channel order for signal configurations was standardized as follows (from top to bottom) (1) flow rate; (2) voided volume; (3) $p_{ves}$; (4) $p_{abd}$; (5) $p_{det}$; (6) infused volume. B: Annotations corresponding to cystometrogram/pressure-flow study tracing.

Neurourology and Urodynamics DOI 10.1002/nau
contractions >10 cmH2O. The annotation was to be placed at the maximum Pmax during the contraction and with indication of whether leakage occurred. Filling continued until the subject reached maximum cystometric capacity (MCC). At this point, the CMG was stopped and the tracing was annotated to identify MCC.

The PFS was conducted after the subject reached MCC. If for any reason the subject had to change positions between the filling and the voiding portion of the assessment, the pressure transducers were repositioned so that they were maintained at the level of the upper edge of the pubic symphysis. However, position changes were only allowed for medical safety reasons and subjects were encouraged to remain in the same position annotated for previous studies. All changes in position and adjustment of transducers were annotated on the tracing. Prior to voiding, the subject was instructed to cough to reassess the dynamic response of the pressure channels. This was annotated as “pre-void cough” and concordance of >70% between Pves and Pabd signals was confirmed. The tester was trained to troubleshoot the system for any discrepancies prior to the void. Since establishment of stable baseline pressures prior to voiding is critical for reliable interpretation of the PFS, testers were instructed to identify a stable baseline pressure (minimum of 5 sec) prior to subjects voiding. A PFS baseline pressure annotation was made after the cough and before giving permission to void. If the subject voided pre-maturely, the PFS was to be repeated. Testers were also instructed not to annotate PFS baseline if DO was present. Instead, testers were instructed to wait for the involuntary detrusor contraction to subside and then reassess if annotation was still appropriate. If the PFS needed to be repeated because of DO or premature voiding, the tester was allowed to refill at a slower infusion rate.

Voiding parameters during the PFS that were assessed included: Qmax, Vcomp, and PdetQmax. In addition, testers were instructed to annotate maximum detrusor pressure (Max Pdet) if this was greater than PdetQmax. After the void, a cough confirmed that the catheters were still functioning and was annotated on the PFS tracing. The 70% concordance requirement was relaxed since this response could be blunted by the effect of an empty bladder on the urethral catheter. Instead, testers were instructed to look for an appropriate deflection in each tracing and if not seen, testers were instructed to irrigate the tubing in the event that the catheter was occluded by the decompressed bladder wall. If catheter displacement occurred, the PFS was to be repeated. A post-PFS PVR was measured via the urethral catheter. Examples of the CMG and PFS including orientation, concordance checks, and proper annotation are provided in Figures 2A,B.

Investigators and urodynamic testers were instructed to review the PFS tracing to determine if minimum validity standards were met. A PFS was defined as invalid if the catheters failed to function properly or became displaced during the first 50% of the voided volume; if the subject leaked more than 100 ml during an involuntary detrusor contraction; cough concordance was <70%; the PFS baseline had not been established for a minimum of 5 sec prior to the initiation of the void; or the tracing was illegible. Once the tracing was deemed valid by the investigator and urodynamic tester, the second PFS was performed in an identical fashion. Both tracings were forwarded to the central reader who made the final decision on the validity of the tracings.

Data Management and Review

All UDS were standardized across sites in accordance with recommendations of the ICS Standardization Committee of Good Urodynamic Practice. A Central Reader was instituted to avoid inter-rater variability. Some standardization procedures were put in place to facilitate uniform interpretation by the Central Reader; these included a standardized channel order for signal configurations, and consistent annotation of all parameters supporting primary and secondary urodynamic objectives of the protocol. All tracings were sent to the Central Reader within 24 hr of testing completion. Tracings were reviewed within 48 hr, at which point sites were notified of any invalid tracings so that a repeat UDS could be performed prior to randomization. If a repeat UDS was not possible, the subject was discontinued from the trial. For valid tracings, the Central Reader calculated bladder outlet obstruction index (BOOI)6,7 and notified the site within 48 hr for stratification purposes. If a tracing was deemed invalid or if a tracing was indicative of inappropriate subject selection, further training of the site personnel to correct the issues was performed. The Central Reader was responsible for interpretation and data entry for all urodynamic parameters to be included in the final analysis, including selection of the more appropriate of the two PFS at baseline and endpoint. In all cases, if the first PFS tracing was deemed valid, these data were selected for analysis purposes; only when the first PFS tracing was invalid was the second tracing chosen for analysis. To ensure data quality, a Secondary Central Reader independently entered UDS results in the electronic database. The Sponsor compared UDS data entries for discrepancies, with queries issued to the Primary Central Reader to correct or confirm data.

Statistical Analysis

In order to assess the test–retest reproducibility of UDS, select PFS and NIF measures were analyzed to determine the correlation coefficients of the placebo group for baseline versus endpoint. PFS measures of PdetQmax, BOOI, and Qmax along with NIF measures of Qmax and PVR were included in this assessment.

As described by Abrams, the standard error of the change from baseline to week 12 in PdetQmax was assumed to be 30 cmH2O. Based on this, in order to have an 80% probability that the confidence interval included a difference of >15 cmH2O in PdetQmax when the true difference was 0, a total sample size of ~190 subjects (95 subjects per treatment group) was needed in order to yield ~128 subjects with both baseline and end-of-study urodynamic results (allowing for up to 30% of subjects to have baseline but not an end-of-study UDS).

For the reproducibility analysis, baseline, endpoint, and change over 12 weeks in the placebo arm were summarized using descriptive statistics, including mean, SD, inter-quartile
range (IQR), and correlation coefficient (r). The number of invalid UDS at baseline and 12 weeks was also examined.

RESULTS

Two hundred men with moderate to severe LUTS (IPSS ≥13) with or without BPO were randomized to receive either tadalafil (N = 99) or placebo (N = 101). Valid baseline and endpoint PFS data were available for 172 subjects (tadalafil N = 83, placebo N = 89). Of the 226 subjects who underwent UDS at baseline, 21 (9.3%) of the tracings were deemed invalid by the Central Reader, while only 0.6% (1/174) of endpoint tracings was considered invalid. Although a total of 173 endpoint tracings were deemed valid by the Central Reader, no data were entered for one of these subjects. Analysis of the change from baseline to endpoint in the primary outcome data (PdetQmax) resulted in a calculated SD of 15.

To determine the test–retest reproducibility of urodynamic measures, a post hoc analysis of select measures from the placebo arm was performed (Table II). The data suggest that test–retest reproducibility was very good for the main parameters assessed during invasive pressure-flow UDS (Table II and Fig. 3A), with correlation coefficients being greater for PFS than NIF parameters (Table II and Fig. 3B). As would be expected, due to placebo effect, correlation coefficients were relatively low for baseline versus endpoint IPSS results (Table II).

DISCUSSION

This report details our efforts to develop and implement a standardized urodynamic protocol in a large, randomized, placebo-controlled, multicenter trial in men with LUTS. Although the ICS has outlined clear recommendations and guidelines for good urodynamic practice, variability between sites can present challenges when interpreting UDS data, making it difficult to incorporate UDS in large, multicenter trials. Schafer et al.2 reported the difficulties experienced during implementation of multicenter urodynamics in the ICS-BPH Study which utilized urodynamic tracings from “less well-known” centers. Tammela et al.7 confirmed that differences among urodynamic laboratories made interpretation of PFS difficult, despite adherence to the ICS recommendations. To optimize the use of UDS in multicenter clinical trials, Lewis and Abrams3 provided a descriptive report which included the recommendation that tracings conform to the ICS recommended standards and quality control measures, as well as a process for centralized review. Nevertheless, Nager et al.10 still encountered problems during the implementation of standardized UDS in a multicenter clinical trial on stress urinary incontinence. These problems were taken into consideration when designing the current urodynamic procedures for this LUTS clinical trial.

Many of the difficulties that Schafer et al.2 experienced were related to inconsistencies between centers with regard to zeroing the pressure lines before or during the study, movement related artifact, and poor concordance in pressure transmission between Pdet and Pves. We avoided these difficulties by requiring regular concordance checks during the CMG to minimize the risk of aberrant pressure transmission errors, and by standardizing and training all testers in the zeroing process. The order and scaling of each pressure tracing was also standardized, thereby limiting the variability in urodynamic tracing appearance and ensuring that the Central Reader had a constant orientation. Schafer described some artifacts which were more difficult to correct, including periodic loss of signal, loss of Pdet catheter during voiding, lack of scaling or zeroing or complete loss of signal. Our guidelines required immediate troubleshooting for these problems and if not corrected resulted in the need to repeat the UDS. By training the UDS testers how to troubleshoot and what could invalidate a study, many of these uncorrectable errors were obviated. The low number of invalid UDS in our trial at baseline and endpoint validates our pre-emptive efforts to achieve consistent and accurate urodynamic data from each site.

Further support that our standardized UDS protocol reduced variability can be surmised from the fact that data from the placebo arm provides evidence that test–retest reproducibility was very good for PFS parameters. Additionally, the SD for the change in PdetQmax was much narrower than anticipated and previously described by Abrams et al.8 (15 vs. 30, respectively). While this may, in part, be explained by differences in the populations, it is also possible that our standardized training reduced the inter-UDS variability, thereby improving consistency. A narrower SD could have direct benefit to future trials using multicenter UDS by allowing for a smaller sample size, while maintaining the same power estimates.

Options for urodynamic interpretation include using either the local investigator or a central reader. While the local investigator is closest to the study being conducted and may be the best interpreter for clinical purposes, use of multiple urodynamic interpreters increases the risk of inter-rater errors and reduced reliability. While utilization of a central reader would eliminate the inter-rater reliability issue, it is expensive and labor intensive, particularly when clinical trials require large subject populations and when multiple urodynamic tracings are performed. Recognizing this dilemma, Nager

<table>
<thead>
<tr>
<th>TABLE II. Mean ± SD, Inter-Quartile Range (IQR), and Correlation Coefficients of Subjects in Placebo Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
</tr>
<tr>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td>PdetQmax (cmH2O)</td>
</tr>
<tr>
<td>BOOI</td>
</tr>
<tr>
<td>Pressure-flow Qmax (ml/sec)</td>
</tr>
<tr>
<td>Free-flow Qmax (ml/sec)</td>
</tr>
<tr>
<td>PVRcath (ml)</td>
</tr>
<tr>
<td>IPSS</td>
</tr>
</tbody>
</table>

BOOI, bladder outlet obstruction index, calculated as PdetQmax − 2Qmax. IPSS, International Prostate Symptom Score; IQR, inter-quartile range; PdetQmax, detrusor pressure at peak urinary flow rate; PVRcath, post-void residual volume measured via catheter; Qmax, peak urinary flow rate; r, correlation coefficient.
et al. utilized local investigators for interpretation and developed interpretation guidelines which all urodynamic interpreters were expected to adhere to when reviewing tracings. In this trial we elected to use a central reader in order to minimize variability in interpretation of UDS tracings. Although we did not conduct a formal analysis of central reader data quality, a secondary central reader was employed to enter all UDS data independently. While several data discrepancies existed, the majority of differences were discovered early in the study and were the result of communication gaps rather than differences in interpretation of tracings.

In spite of our standardization efforts, there were a number of unforeseen issues that arose. The fact that our rate of invalid studies continued to improve between baseline and endpoint implies that a learning curve was still present. It is not clear whether additional training or practice would have lowered that number of invalid studies in the early part of the trial. When additional sites were added, it was difficult to provide the centralized training on an individual basis in a cost-effective manner. Perhaps the most important lesson was that inter-site reliability could have been optimized if all sites were required to use the same urodynamic equipment and software. This would have allowed for tracings from all sites to have identical orientation and appearances and it would have allowed for construction of an electronic central repository which could then be utilized for quality assurance and for central interpretation, as advocated by Nager et al.

**CONCLUSION**

This report is intended to augment the recommendations and guidelines set forth by the ICS, in order to improve consistency across multiple UDS testers in a multisite clinical trial. These data indicate that by incorporating standardized training and a central interpreter, test–retest reproducibility over a 12-week study is very good for a large-scale clinical
trial, with correlation coefficients being greater for PFS than NIF parameters.

While the circumstances and priorities of other urodynamic investigators will vary and may require variations on these urodynamic procedures, we hope that our experience with the development of our centralized urodynamics protocol can serve to help other investigators wishing to use urodynamics in multicenter clinical trials for men reporting LUTS with or without evidence of BPO.

ACKNOWLEDGMENTS

Michael Anne Shane, MS, assisted with the preparation of this manuscript. This study was funded by Eli Lilly and Company.

REFERENCES