**National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)**
**National Institutes of Health (NIH)**
**Meeting on Measurement of Urinary Symptoms (MOMUS)**

**November 14 – 15, 2011**
**Natcher Conference Center, NIH**
**Bethesda, MD**

**Summary Report**

## MONDAY, NOVEMBER 14, 2011

### INTRODUCTION

*Ziya Kirkali, M.D., Senior Scientific Advisor, Division of Kidney, Urologic, and Hematologic Diseases, NIDDK, NIH*

Dr. Kirkali welcomed participants and commented that two scheduled speakers were not in attendance, but there would be replacement speakers.

*Griffin Rodgers, M.D., M.A.C.P., Director, NIDDK, NIH*

Dr. Rodgers greeted conference attendees and stated that lower urinary tract symptoms (LUTS) occur in both males and females, and symptoms increase with advancing age. One half of men and women suffer from urinary storage symptoms and one quarter from voiding symptoms. According to the NIDDK funded project, Urologic Disease in America (UDA; http://kidney.niddk.nih.gov/Statistics/UDA/index.aspx ), in 2000, diseases related to the prostate cost $2.4 billion. This estimate was restricted to outpatient costs and therefore, is likely an underestimate. The U.S. population is aging, which is likely to increase LUTS costs over time.

The American Urological Association (AUA) symptom score has been used for 20 years to measure LUTS, but its usefulness in clinical settings is sometimes disputable. Assessment of chronic disease symptoms is difficult because symptoms wax and wane; understanding these symptom flares is imperative to comprehending the natural history of the disease. Valid measurements used to assess LUTS are essential to this insight process.

The MOMUS meeting is comprised of participants from industry, academia, clinical settings, and government, as well as LUTS patients. This gathering of individuals from various expertise areas and with diverse experiences will greatly assist in moving the field of LUTS measurement and patient phenotyping forward.

*Janine Clayton, M.D., Acting Director, Office of Research on Women's Health (ORWH), NIH*

Dr. Clayton welcomed participants. A recent National Health and Nutrition Examination Survey (NHANES), defining incontinence as leakage, found the prevalence of incontinence to be 51 percent in women and 14 percent in men. Women experience this elevated prevalence because of childbirth. Inter- and multidisciplinary approaches are necessary to understand LUTS, and the

variety of scientists attending MOMUS will help in this endeavor. Dr. Clayton acknowledged the MOMUS planning committee, speakers, patients, and other attendees.

## I. PUBLIC HEALTH IMPORTANCE OF MEASURING SYMPTOMATIC LOWER URINARY TRACT DYSFUNCTION (LUTD)

*Moderator: Kevin McVary, M.D., Professor, Department of Urology, Feinberg School of Medicine, Northwestern University, Chicago, IL*

### We Are All Happy, Why Bother?

*Marcus Drake, M.D., Ph.D., Professor, Department of Urology, Bristol Urological Institute, Southmead Hospital, University of Bristol, United Kingdom*

Dr. Drake replaced Dr. Paul Abrams (Professor, Department of Urology, Bristol Urological Institute, Southmead Hospital, United Kingdom).

There is a wide range of lower urinary tract dysfunction (LUTD) research areas; and with effective communication and general adoption, consensus on assessment can be achieved. Dr. Drake and Dr. Abrams are members of the Advisory Board for the International Consultation on Incontinence Modular Questionnaire (ICIQ). ICIQ has developed a range of tools encompassing LUTS, including male and female urinary symptoms, and vaginal and bowel symptoms. Future initiatives include a urinary diary and other modules in various stages of development (e.g., a module related to catheterization). These modules have been aptly developed in areas of clinical need, and some of the tools have been translated into non-English languages. Researchers, doctors, and patients require these tools and would benefit from a menu where they can identify the appropriate tool to use, based on the question(s) that they want answered. Using the ICIQ questionnaires, symptoms can be catalogued and quality of life (QOL) impacts can be comprehended. The patient perspective, which is needed for prioritization of treatment(s), is of the utmost importance. Simple biological observations may not be the most pertinent to patients and attainment of the desired outcome.

Dr. Drake emphasized the importance of appropriately using terminology. The descriptive term "lower urinary tract symptoms" (LUTS) was coined in the *British Medical Journal* in 1994 by Abrams. LUTS avoids inferring mechanism, as is implicit in some terms, such as the now-discredited term "prostatism". Clinicians also need to be exact in use of recognized terms such as benign prostatic hyperplasia (BPH), benign prostatic enlargement (BPE), and benign prostate obstruction (BPO). "Hyperplasia" is the appropriate term where histological confirmation of pathological changes has been obtained from prostate tissue, contrasting with "enlargement", which can be assessed by digital examination. The term "BPO" should only be used when clinicians have confirmed the presence of such (via cystometry). The U.S. Food and Drug Administration (FDA) maintains the use of the "BPH" term in clinical settings, which is inappropriate and should be reviewed. "Overactive bladder" is also a name open to criticism; the diagnosis may be derived when patients report urinary urgency, and since "bladder" appears in the condition, clinicians may neglect to evaluate all potentially-contributory facets- such as urethra, prostate, vagina, and rectum.

Clinicians hope to achieve diagnoses with questionnaires, but this currently appears unrealistic. Accordingly, naming of symptom assessment tools should be undertaken judiciously. For example, women can score highly on the AUA international prostate symptom score (IPSS), even though the prostate gland is a male genital organ. Dr Drake considered that they record symptoms, bother, QOL, and can indicate natural history changes. But it is vital to ensure that the nomenclature used on these questionnaires is correct, and must be robust and relevant. Because the IPSS omits incontinence, the most bothersome symptom for many men, Dr. Drake suggested it fails to capture key aspects of LUTS. Dr. Drake gave an example of a male patient with LUTS comprising severe bothersome nocturia, incontinence, and post micturition dribble; since none of these are well captured by IPSS, his IPSS and QOL score may fail to reflect a representative insight into his LUTS. These parameters all are captured in the ICIQ-male lower urinary tract symptoms (ICIQ-MLUTS) module, a tool designed to evaluate the full spectrum of symptom severity and bother. An advantage of ICIQ-MLUTS over the IPSS is the assessment of the bother of each separate LUTS, since an overall QOL score fails to reflect the symptom of greatest concern for the individual patient. Dr. Drake thought that combining questionnaires from the ICIQ system enables clinicians to achieve a coherent and comprehensive evaluation relevant in modern-day multidisciplinary practice. For example, urinary and bowel symptom tools should both be used in patients with mixed (urinary and fecal) incontinence.

Only observing symptoms is not adequate for a mechanistic diagnosis. In men and women contributory pathophysiology for urinary symptoms can include storage overactivity, voiding underactivity, outlet obstruction, fluid balance problems, and sphincter weakness. Full evaluation of the situation for an individual is needed to clarify mechanisms. For example, nocturia is not necessarily merely a LUTS; it can be an expression of systemic problems, including endocrine or cardiovascular abnormality.

Dr. Drake concluded with a proposal that tools are already available or in development for full evaluation of the entire clinical picture in LUTS, with the possible exception of patient adaptations for symptom amelioration. New instruments are not necessary for every new trial. The ICIQ can already serve as a platform for patient evaluations and research, but that the professions have failed to deploy them to their potential—perhaps in part due to regulatory and guidance deficiencies. A new generic tool for LUTD assessment cannot be justified where such effective tools are available, along with the platform to develop additional items according to identified need.

**The Public Health Impact of LUTD Symptoms**

*Mark Litwin, M.D., M.P.H., Professor and Chair, Department of Urology, David Greffen School of Medicine at University of California, Los Angeles, CA*

Dr. Litwin began by sharing information on the UDA compendium publication (http://udaonline.net ).
To share recent UDA findings, Dr. Litwin explained that UDA was intended to document the burden of illness in practice patterns, epidemiology, diffusions and adoptions of new technologies, costs, access, and quality of care (QOC). The project then assesses the outcome of

care. This is achieved by gathering and analyzing epidemiological and insurance claims data. The results obtained will inform public policy and determine promising areas for future research.

The burden of illness is defined by the prevalence and incidence, length of hospital stays, frequency of physician's office visits, emergency room (ER) and ambulatory visits, nursing home admissions, and direct and indirect costs. Unfortunately, QOL data cannot be captured from insurance claims data, and so it is necessary to communicate with patients directly.

Billions of dollars are spent annually on urinary tract infections (UTI) in women ($2.474 billion) and men ($1.1 billion). Incontinence in women has an annual price tag of $500,000. In Medicare expenditures, urinary symptoms account for significant amounts of money, which increases with patient age.

Another cost of burden is physician's office visits. There are millions of such visits annually for primary or secondary diagnoses for various urinary conditions. Dr. Litwin hoped to impress upon participants the time, effort, and financial resources spent on patients with benign urinary tract symptoms.

Outpatient visits for BPH are increasing and the use of inpatient visits, hospitalizations, and inpatient procedures (e.g., transurethral resection of the prostate [TURP]) are decreasing. There has been an influx of minimally-invasive BPH treatments for men. The direct cost of BPH is $1 billion, with 4.5 million annual visits to physician's offices for primary diagnoses of BPH.

Men with prostatitis visit physicians who often cannot offer them effective treatments. Prostatitis is largely an outpatient condition, and a high number of men with chronic prostatitis symptoms undergo procedures that are generally understood to be ineffective.

Interstitial cystitis (IC) and painful bladder symptoms are the most common LUTS in women, although annual office visits have decreased. Many women with IC see a urologist for treatment, although many see their primary care physician or gynecologist. The cost of IC is not significant for women or men, with expenditures ranging from $7,000 to over $25,000 annually.

Urinary incontinence is highly preventable; however, 10 to 15 percent of women have daily leakage, which is a significant burden. Incontinence in men is not as severe, but leakage increases in prevalence as men age.

The UDA project researchers also examined UTI in women and focused on racial variations in both symptoms experienced and healthcare accessed. In women, various ethnic groups accessed care for UTIs differently (e.g., physician office visits, ER visits, and so forth). This finding lends itself to questioning if the disease is expressed differently in women of varying ethnicities. There is a trend towards outpatient care for UTI, even in treatments that are traditionally inpatient. Not taking into account medications, the annual cost for UTIs is $2.474 billion for women.

**Intervention for Symptoms: Towards Outcomes That Matter**
*William Lawrence, M.D., Medical Officer, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality (AHRQ), Rockville, MD*

Dr. Lawrence, an end-user of MOMUS-type results, discussed the reasons for AHRQ interest in MOMUS. The AHRQ mission is "to improve the quality, safety, efficiency, and effectiveness of health care for all Americans".

Comparative effectiveness research (CER) is the conduct and synthesis of research that compares the benefits and harms of different interventions and strategies to prevent, diagnose, and treat minor health conditions in "real world" conditions. The purpose of CER is to improve health outcomes by developing and promulgating evidence-based information to inform decision making. To provide this information, CER must assess a comprehensive array of health-related outcomes for diverse patient populations and subgroups. AHRQ endeavors to provide evidence that assists people in making decisions about their care.

The Patient-Centered Outcomes Research Institute (PCORI) has definitions of patient-centered outcomes research (PCOR) that encompass the concept of research answering patient-focused questions (e.g., How can I improve outcomes that are important to me? How can the health care system improve my chances of achieving the outcomes I prefer?). To do this, PCOR assesses the benefits and harms of procedures and examines outcomes that people notice (e.g., survival, function, and QOL).

CER/PCOR aim to assist in decision making and require outcomes that are important to patients. Symptoms are important to study and understand, but they may not be sufficient. Patient preferences, impact on function, and QOL must be considered. As an example, Dr. Lawrence described a patient who had prostate cancer, LUTS, as well as bowel problems from the radiation cancer treatment. Due to his career requiring extensive vehicular travel, what would be fecal urgency in an office setting with easy access to a restroom became fecal incontinence during his business travel. This was devastating for the patient, and an illustration that QOL and impact on individuals should be contemplated in addition to measuring symptoms.

The AHRQ effective health care (EHC) program has the CER goal of improving the effectiveness and efficiency of healthcare delivered through Medicaid and children's programs by focusing on what is known, research gaps, and clinical effectiveness. They work to be attentive to underrepresented subgroups, comorbidities, and treatment heterogeneities.

As a framework for the CER approach, the EHC program is maintaining a vision of the future. The program is mainly evidence-based and encompasses the syncretism of this evidence. Literature is synthesized to determine what is known and what fits into research gaps. A main interest is to translate research into meaningful and relevant messages for patients and clinicians and to ask key questions (e.g., What constitutes an adequate diagnostic evaluation for women in the primary care setting on which to base treatment of urinary incontinence [UI]?).

The EHC program does not focus on life extension, but instead focuses on creating an increased QOL. Dr. Lawrence concluded by reminding participants to consider a broad array of symptoms and contemplate function, QOL, and outcomes. More information about the AHRQ EHC program can be found at the following URL: www.effectivehealthcare.ahrq.gov.

**Measurement of PROs: The PROMIS Initiative**
*William Riley, Ph.D., Program Director, Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute (NHLBI), NIH, Bethesda, MD*

Dr. Riley stated that NIH has funded the Patient Reported Outcomes Measurement Information System (PROMIS) for many years with the initiative to create harmony across studies and disease entities. Another goal is to develop and evaluate patient-reported outcome (PRO) item banks for a range of chronic disease and outcomes research.

PROMIS began in 2005 and continued through 2009 as PROMIS I, in which measures were developed. PROMIS II (2010 to 2013) works to validate the measures developed in PROMIS I and create new measures in areas with gaps. To date, PROMIS has involved over 40,000 patients, resulted in over 100 publications, and released 30 item banks (e.g., pain, fatigue, sexual function). Web-based administration and scoring systems are available.

Dr. Riley described how one measure being developed can assist PROMIS researchers in achieving data harmonization. Previously, the prevailing concept was to do "best in class" consensus measures and assume that it will be used appropriately and willingly by everyone. This approach is potentially limiting because it requires adoption by the research and clinical communities and does not harmonize with previous data; it is unclear how to integrate future measurement advances.

An alternative to the above-described methodology is to co-calibrate all of the data to a single scale or metric. This would allow tools to be treated separate from their original scales and be interpretable, actionable, and flexible. In this way, data can be harmonized with previous research and adapted as measurement science advances. Dr. Riley exemplified this by explaining that items can become obsolete over time and using item-response theory (IRT) can overcome this problem. For example, the measurement of blood pressure previously required the use of mercury sphygmomanometers; however, the current sphygmomanometers do not use mercury and measure blood pressure differently. With the same scale, the new measurement method can be utilized and data are easily compared to measurements taken from earlier sphygmomanometers.

For IRT, a category response curve can be constructed. Computerized adaptive testing allows subsequent question items to be based off of previous questions and responses. This permits a fairly precise estimate of a patient's condition. IRT also allows for weighted responses because not all symptoms carry the same level of importance. For example, a patient response that life is no longer worth living is more severe and should be weighted differently than a patient response that s/he gets upset.

Certain item banks (e.g., fatigue) can be used in many different trials and for many different diseases, such as chemotherapy trials, diabetes, Parkinson's disease, and so forth. Measurement precision (standard error) indicates that measures function optimally when patients are in the mid-range for any given measurement, with less optimal precision obtained when patients fall into distribution "tails" or "extremes."

The PROMIS assessment center (http://www.assessmentcenter.net/) is web based and allows researchers to register and gain access to the PROMIS tools free of charge. Researchers can select any measure that they have interest in from multiple studies, in many forms, and can build in additional scales. At the end of a study, researchers can track how patients are doing and responding, and export the data in any one of various formats. For further information, the main PROMIS website (http://www.nihpromis.org/) contains information about all of the available item banks and answers related questions.

**PRO in the Evaluation of Medical Products for Regulatory Approval**
*Laurie Burke, R.Ph., M.P.H., Associate Director for Study Endpoints and Labeling, Office of New Drugs, Center for Drug Evaluation and Research, FDA, Silver Spring, MD*

Ms. Burke presented a summary of the FDA clinical outcome assessment (COA) workshop, held on October 19, 2011, and paid special attention to COA development.

The goal of COA measurement is to assess the impact of treatment on how patients feel and function in their daily lives in order to support conclusions concerning treatment benefit. In addition to survival, COAs measure treatment benefit either directly (e.g., symptoms) or indirectly (e.g., biomarkers) and can represent treatment effectiveness or comparative safety. Indirect assessment requires empirical justification for the value as a replacement for how patients survive, feel, or function. There also are several types of outcome assessments including reported, non-reported (biomarkers), and survival.

COAs are critical to understand the benefits and harms of treatments. Rigorous development leading to well-defined and reliable COAs in the clinical trial context is required before they can be deemed adequate to support FDA's substantial evidence standard by use in adequate and well controlled studies. The FDA reviews COAs when results from primary or secondary endpoints are represented in statements of drug effectiveness in labeling.

There is a new qualification process for PRO assessments that is available at the following FDA website: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. This document is in draft form and is still undergoing revision and finalization. An FDA goal is to have publically available assessments that will circumvent repeated review and advisements given on diseases that are similar across drug development. Qualification decisions represent the conclusion that assessment results can be relied upon to measure a specific concept and have a specific interpretation and application, within a specific context of use. There are two FDA review processes for COA review: drug application and qualification, with the requirement that any conclusion of treatment benefit be based on substantial evidence.

The FDA guidance document defines instrument development by beginning with a definition of the "context of use." The manner and purpose of usage of a COA includes the targeted population, type of trial, and "other" (e.g., drug mode of action, instrument format, onset of action, mode of administration). An endpoint model is a simple diagram or description of the planned clinical trials explaining what the primary and secondary endpoints will be.

The assessment of treatment benefit can consist of direct and indirect evidence. Direct measures include how a patient feels and functions in daily life and can be reported by the patient, an observer, or a clinician. Indirect measures can include the same reporting format, but include biomarker measures. While biomarkers are valuable in supporting conclusions of treatment benefit, additional evidence is necessary for full FDA approval. There is a continuum of indirect measures because not all are equivalent regarding replacement value and proximity to direct measurements (e.g., although an indirect measure, an exercise test is closer to a direct measurement than body temperature). Direct measures also have a continuum. Interpretations of trial results depend on understanding how a treatment impacts the core disease-defining concepts first.

The FDA also emphasizes the need for endpoints that measure the concept represented in targeted claims. For example, an advertisement claiming "qualify of life significantly improved" cannot be used or implied if this was not measured directly and exactly.

The conceptual framework of an instrument is another tool that can be used to evaluate if an instrument is adequate to meet study objectives. Using this tool, the concepts being measured can be explicitly stated to provide an understanding of what an instrument measures and what is important in the "context of use" under review.

When concepts and a context of use have been defined, the FDA reviews whether content validity is established for the new or revised COA. Empiric evidence is required that the instrument measures the targeted concept in the appropriate context of use. Content validity must be established prior to interpretation of construct validity, reliability, or sensitivity. Methods to institute content validity are iterative, including a literature review, expert opinion, qualitative research, and quantitative analyses. Quantitative methods to document content validity can be used iteratively and graphical displays can illustrate if a measurement ruler has been created successfully. The overlap between response options and the uniqueness of information provided by response options also can be evaluated. High-quality cognitive debriefing of the final instrument confirms content validity.

Rigorous COA development can minimize many types of variance (e.g., heterogeneity, nonrandom error, random error, experiment error). Trial variability and increased error amplifies the need for a large clinical trial size. The timeline for COA development should include defining the context of use, hypothesizing concepts, finalizing the instrument content, assessing respondent understanding, establishing other measurement properties, and preparing a COA dossier, with the final goal of creating a well-defined and reliable COA.

Regarding the topic of validation, it is incorrect to speak of a "reliable and valid" test because they are not immutable, inherent properties, but rather they are results of an interaction between the COA and the context of use.

The FDA has an enforcement program and notices of violation or warning letters result from advertisements with unsubstantiated claims.

As new technologies become available, the FDA is committed to updating their guidance documents to reflect changes in policy, working with stakeholders to make qualified COAs publically available, and to conduct internal tracking for COA reviews. FDA involvement in the discussion on good measurement is intended to drive the development of better information for patients. Because patients should know how new treatments will impact them, attention to COA development at the earliest stages of drug development will ensure that better information is available to patients about their medical treatments.

***Discussion***

Dr. Gerald Timm (GT Urological, LLC) questioned why there was no mention of device or objective assessments. Ms. Burke indicated that the focus of this session is symptoms. However, the PRO guidance presents standards for good measurement science that also apply to all outcome assessment in clinical trials. Responding to another query, she explained that a patient diary can be used to capture many types of outcomes—symptoms, impact of symptoms on functioning or emotions, activities, or events. The key is to first identify the important thing to measure in the targeted patient population and then to identify the best way to measure that thing. For symptoms, a PRO assessment is required. For physiologic function, a clinical test or observation may be the best option. The determination of what to measure requires clinical determination of how to define the population under study and how to define improvement in their disease or condition. Reliability and validity are not immutable; a patient may say in a diary that they are "cured," but clinical tests could show no difference in their incontinence. The change that occurred may be in the patient's perception of urgency. This is an example of why it is important to define the disease and improvement in the disease before the outcome assessment is defined. Clinically, will treatment benefit be concluded if patient perception of improvement alone is demonstrated? Or does treatment benefit depend on a clinician opinion regarding patient improvement?

Dr. Robert Star (NIH) questioned if the FDA COA workshop (October 19, 2011) had been captured electronically. Ms. Burke responded that the transcript and PowerPoint slides will be made available. Dr. Star emphasized the importance of that workshop in discussing the critical concepts of context of use, diagnosis, and outcome. The intended use of instruments is critical for appropriate outcome measures. Dr. Lawrence commented that in observing the continuum from symptoms to impact on QOL, there are social, cultural, and personal factors that influence how patients adapt to symptoms. There is a tension between physiological impairment and symptoms/QOL; doctors tend to be interested in the former, although patients are interested in the latter.

Ms. Jane Meijlink (International Painful Bladder Foundation) commended the patient-centeredness in the United States and indicated that this is lacking in Europe. Dr. Drake concurred but indicated that there is increasing responsiveness to this issue from the medical community.

Dr. John Kusek (NIH) stated that, at the intersection between FDA approval processes and promises, there are many indices in the urology field and some people argue that there should be outcomes in drug approval. Dr. Lawrence indicated that there is an interagency working group on clinical outcomes. PROMIS was intended to be a generic tool encompassing many diseases.

Patient groups exist that do not fit into a disease paradigm, and this is challenging to researchers regarding progression and change in disease. Measurement sensitivity, therefore, is imperative. Additionally, disease pathology does not begin and end when measurements are collected and so the natural history and fluctuations of diseases are not being captured effectively.

Dr. Claus Roehrborn (University of Texas) thought that Dr. Drake's critique of incorrect terminology may have overlooked the fact that drug development has FDA approval as the intended end product. When the FDA is aware only of one word, BPH, is that not stifling? Another participant commented that regarding language use, the FDA mirrors what is in the current scientific literature. If the urology community changes terminology, the FDA will follow suit.

PRO instruments are not diagnostic—they are outcome assessments. The adequacy of a PRO instrument, however, depends on the context of use that includes the full set of inclusion and exclusion criteria for the clinical trial including all diagnostic assessments. Questionnaires can be helpful in measuring incontinence, but there are multiple ideologies and diagnoses that researchers must be mindful of and put in the context of the research question. Dr. William Steers (University of Virginia) commented that often before a device or drug can go to trial, a signal must be observed in an animal model. He expressed concern whether there are any efforts to translate these other systems to humans (e.g., does pain and urinary frequency in animal models translate to humans?). Dr. Drake agreed that these points about the reliability of animal models, as well as FDA terminology, are fundamental.

## II. PATIENT-FOCUSED APPROACH TO AN INVISIBLE CONDITION

*Moderator: Lisa Begg, Dr.P.H., Director of Research Programs, ORWH, NIH*

Dr. Begg stated that this session will focus on patient perspectives.

### What I Expected and Found After Prostatectomy

*Robert Wedgworth, M.S., University Librarian and Professor Emeritus, University Library, University of Illinois at Urbana-Champaign*

Mr. Wedgworth, a 74-year-old man who lives in Chicago, IL, was diagnosed with BPH in 1995. The symptoms worsened gradually until 2010, during an Australian holiday, when he was unable

to urinate for 12 hours. He visited a clinic and found relief via a catheter. When he returned to the United States, he was told that because that incident occurred once, it may occur again. In response to this, he underwent a prostatectomy in February 2011 and has found a significant increase in QOL. He explained that he now can sit through an entire movie or a one-act play without discomfort; he has complete evacuation 90 percent of the time and minimal leakage (10% of the time).

The recovery time estimate he received for the prostatectomy was 8 to 10 weeks, but it was 12 weeks before Mr. Wedgworth began to feel better. He is able to achieve erection and sexual satisfaction without ejaculation and has no pain.

The psychological dimensions of BPH outweigh the physical symptoms. He expressed great anxiety at previously having to identify bathroom locations in every building, wondering if he would make it to the bathroom in time, handling long meetings, and feeling that with BPH he was in a "waiting line" for developing cancer. As the symptoms of BPH worsen, the tribulations that patients endure increase. For example, following an extended flight from South Africa to England, Mr. Wedgworth had to run to the bathroom following deplaning. There were 20 people waiting in line before him and he was forced to beg a private lounge guard to let him use that bathroom instead. These anxieties are not given the consideration by doctors and researchers that they deserve. These concerns are what weigh on patients' minds.

Mr. Wedgworth expressed gratitude for the assistance of his wife. A 4-week automobile trip through New Zealand was relaxing and lacked anxiety because, at his wife's suggestion, he wore an incontinence pad. He was no longer required to spend mental time worrying about bathroom locations and the timing of restroom stops.

**I Expect to Urinate Frequently, but This Pain is Wearing Me Out**
*Laura Santurri, M.P.H., Chagrin Falls, OH*

Ms. Santurri stated that she experiences IC, otherwise known as "painful bladder syndrome." This diagnosis was one of exclusion and there are no real treatments. The prevalence of IC has been increasing and now afflicts 3 to 8 million individuals in the United States, most of these being women. There is no evidence that the condition is psychosomatic; however, stress can exacerbate symptoms. Symptoms vary greatly.

Ms. Santurri is a 29-year-old woman from Ohio and has been afflicted with IC since 1997. After presenting with urinary pain, she first was diagnosed with a UTI and prescribed antibiotics. Although the infection disappeared, the symptoms did not. After 4 years, she finally received the IC diagnosis and felt relief with the knowledge that her pain and bother were real and not conjured. The IC has become more painful with time and, although she has found an effective treatment, the symptoms still wax and wane. She has developed a list of comorbidities, including irritable bowel syndrome (IBS).

Although IC symptoms vary by patient, Ms. Santurri's primary symptoms include frequent urination and the need to relieve herself immediately. One member of her support group

reported needing to urinate 60 to 70 times a day, which is of course, highly disruptive to work, travel, and a social life. This particular individual would not leave her house because of IC. Ms. Santurri explained that IC patients cannot simply "hold it" because the condition is quite painful. The pain can radiate to include significant abdominal pain and back pain. After being forced to painfully wait to urinate, IC patients often cannot urinate at all, which increases the pain.

The need for frequent urination can be highly disruptive to sleep because of the recurrent need to wake up and urinate or due to the pain symptoms. The pain that is experienced by patients can be acute or chronic, and can occur in a variety of forms, including general pelvic pain, upper thigh pain, sharp pain during urination (e.g., burning or feeling like small razors are being excreted), and pain during sexual activity. This latter pain symptom encompasses arousal during and after a sexual act. Ms. Santurri expressed that this symptom was the most bothersome for her and affected the intimacy of her former marriage, as she could not have sex more than five times annually due to the intense pain.

Specific symptoms and severity vary by patient and it is important for clinicians and researchers to understand what creates the greatest bother in patients. Symptom indices often do not adequately measure the pain level. Measuring pain is key, as well as QOL. She stated that although the disease is incurable, if she can be functional then she will be able to cope effectively.

**Am I Aged Enough to Develop Urinary Symptoms? Wait; and Pain?**
*Thomas Colclasure, Prostatitis Foundation*

Representing the Prostatitis Foundation (www.prostatitis.org), with a mission statement that calls for the education of the general public about prostatitis and support of NIH via grassroots funding and newsletters, Mr. Colclasure narrated several patient stories.

The first patient, a 19-year-old man, began experiencing symptoms at age 14. Originally, there was slight pain during urination, but after several examinations by a doctor, it was concluded that there was nothing amiss. After visiting a second doctor, a rectal exam indicated prostatitis and when prescribed an antibiotic, the pain subsided. However, the pain persists when he is not taking an antibiotic, and the patient was forced to quit school due to the inability to sit for long periods of time.

Another patient who is 48 years old experiences intense pain, and has been examined by many facilities. Painful symptoms first appeared on long truck rides. All therapies that have been attempted (e.g., physical, herbal) have been ineffective and the patient questions how long he must live with so much pain.

A third patient, 54 years old, has been handling prostatitis for many years. His first indication occurred during a routine physical examination when the doctor told him that his right lobe was enlarged. Following an examination 2 years later, he felt pressure in the prostate that did not subside and was diagnosed with prostatitis. The drugs he had been given were ineffective, even after visiting 12 specialists. Only one specialist examined his prostate fluid. This patient was

mentally and physically fatigued to the extent that he could no longer work, and other normal activities were not possible. Prostate massages and warm water baths improved his condition and his prostate is now normally sized. It is the opinion of the Prostatitis Foundation that profound efforts are needed to solve this disease.

**Incontinence: From Beginning to End?**
*Harry "Doug" Swank, Boyds, MD*

Mr. Swank spent his entire career in the federal government, 33 of those years at NIH. He was diagnosed with prostate cancer at the age of 55.

A prostatectomy was performed in 1996, which led to incontinence, a life altering situation. He suffers from frequent urination during the day and night and attempts to limit fluid intake to combat this; however, decreased fluid consumption causes constipation and decreased bowel movements. Urinary leakage at night is not helped by pads because they are ineffective when a person sleeps on their side. Due to nocturnal "accidents," he takes showers in the middle of the night and must get his clothing and bedding reorganized, which is disruptive to sleep.

Sexual intimacy also suffers because leakage during sexual activity is undesirable. Mr. Swank's wife is understanding, however. His motto is to not waste today worrying about tomorrow.

Mr. Swank has learned to maintain a supply of Depends® products in all locations where he lives (e.g., home, work, trailer, motor home, pant pockets, vacation home, and car). His surgeon suggested Kegel exercises, but Mr. Swank found them ineffective. He considered using a clamp for the incontinence, but opted against this option. He underwent surgery by a urologist to have collagen injected near his urethra; however, this failed to decrease his symptoms. An incontinence expert later repeated the collagen injection with positive results. The decrease in symptoms lasted for 2 years before some collagen dispersed and his incontinence began again.

After 5 more years, Mr. Swank's prostate cancer returned, but the specific location of this cancer could not be determined. He underwent radiation therapy on his prostate bed, but 6 months later his prostate-specific antigen (PSA) levels were severely elevated. A collagen repeat treatment was not effective to stop the leakage and he constructed a clothes pin-like device to stop his incontinence. Any such device, however, is uncomfortable.

In 2009, Mr. Swank opted for a urinary bladder sphincter prosthesis which halted leakage for a couple of years. Leakage has slowly begun to return, and he still wears Depends® for mental reassurance. The prosthesis currently requires repressurization; however, the leakage is minor and he is happy that he may not undergo this procedure due to concerns about the prosthesis eroding through his urethra.

Mr. Swank expressed dismay at the insistence of clinicians to weigh his urinary incontinence pads because the amount of leakage depends on innumerable factors (e.g., daily activity, stress, weather). He currently is involved in a trial for advanced prostate cancer and is open to helping others.

***Discussion***

There currently is an NIH study observing IC and other organ diseases, and it examines IBS and other comorbidities.

Mr. Swank commented that the artificial sphincter can only be moved a couple times before it will erode through the urethra and so it may be a positive thing that he did not have the procedure done when he was younger, as the prosthesis could already be rendered ineffective at his current age. A conference participant was identified as the developer of the artificial sphincter.

It was mentioned that the necessity of weighing urinary incontinence pads is due to it being a clear incontinence indicator and also can determine the level of incontinence that a patient suffers from. The level of bother can be used to measure incontinence, but this depends upon individual patient personality and mentality; however, weighing pads is a direct incontinence measure.

Mr. Wedgworth commented that he likely would have had a prostatectomy performed earlier had he known what he currently knows. Clinicians consistently informed him of his large prostate size, but he did not understand the unusualness of his conddition. The prostate size exacerbated the frequency of his conditions and he would have undergone TURP earlier if he had known.

Ms. Santurri responded to a question by stating that by urinary "urgency," she means pain. She has no fear of leakage, but only the extreme discomfort of burning pelvic pain, abdominal, and back pain.

Dr. Drake commented that the medical profession does not listen to patients well. He thought that part of the treatment for all of the patients in the panel was to finally converse with a doctor who listened and comprehended. He lamented the need in the medical field for patients to be in and out of the consulting room in 10 minutes. He questioned how the medical community can meet patient needs appropriately. Mr. Wedgworth added that the questionnaires that patients are required to fill out are different from manual examination results, and how these both relate to what patients are experiencing can become convoluted. Ms. Santurri concurred and stated that the AUA treatment for IC begins with education, stress management, and self care, before moving into oral medications and therapies. This can create a challenge because urologists may not want to treat patients medically, and in fact have little experience in treating patients with chronic conditions, such as IC.

Regarding location of patient-friendly information, Mr. Wedgworth commented that due to his career as an academic librarian, he was able to find the information he required. He stated that websites, such as WebMD, can be helpful to patients because they are consistent, simple, and straight-forward.

Responding to whether questionnaires appropriately capture symptoms that are important to patients, Mr. Swank expressed no issue with the questionnaires because he simply did what his

doctor prescribed. Ms. Santurri informed the participants that there were problems surrounding her ability to communicate when she was having flares; however, after having IC for 14 years, she has grown accustomed to it and knows the appropriate time to do self care versus visit her doctor. Mr. Wedgworth explained that he headed an adult literacy foundation for several years, and the optimal way to measure the level of an adult's literacy and assess the appropriate literacy program, is to listen to their individual story. Exactly what the person is interested in acquiring from the program can be learned by doing this. He compared literacy to urinary symptoms; different patients have different wants and needs. Mr. Swank added that on a questionnaire he took at a research hospital, pain during ejaculation was not listed because the doctor did not feel this was a symptom; but Mr. Swank felt it was relevant.

## III. WHAT IS CURRENTLY MISSING IN THE MEASUREMENT OF LUTD SYMPTOMS?

*Moderator: Gopal Badlani, M.D., Professor, Department of Urology, University of San Diego, California Health System*

### Current Instruments to Measure Symptomatic LUTD for Diagnostic, Statistic, and Outcome Purposes

*Gopal Badlani, M.D., Professor, Department of Urology, Wake Forest University ,Winston Salem, NC.*

Dr. Badlani began by stating that his views, and not that of the AUA, are being presented. He discussed the currently available instruments for LUTD, and commented that these instruments overlap with associated BPH and erectile dysfunction (ED) diseases, as well as prolapse in women and sexual dysfunction.

When considering measuring instruments, it is important to identify the intended user (e.g., a primary care physician [PCP], gynecologist, urologist, urogynecologist, patient) and the appropriate level of questions regarding literacy. He also indicated the necessity of identifying the intended use of the questionnaire (e.g., diagnoses, following patients, outcome measurements, research).

An example questionnaire may query the frequency of urinating during the day and night, but this is not the complete story. Solely observing the total questionnaire score eliminates the details of urinary volume, urgency, and dysuria, for example.

First developed and published in 1992, and subsequently adopted by the World Health Organization (WHO), the IPSS is the most commonly used symptom score globally. It has been translated into more than 40 languages and despite its flaws, the wide use of it makes it valuable. There is significant re-test variation, meaning that one application may not be sufficient and long-term use is necessary.

The AUA symptom score (AUASS) is gender specific to males with LUTS. It is not intended to diagnose LUTS due to BPH/BPO, but has been useful in clinical research. Dr. Badlani

emphasized that outcome measures must be multidimensional because there is not a single test that can fully gauge effectiveness.

The Danish prostate symptom score (DAN-PSS) is intended to assess LUTS, evaluate treatment options for uncomplicated BPH, and predict bladder outlet obstruction (BOO). While not predicting BOO, it does as well or better than the AUASS.

The short-form Urogenital Distress Inventory (UDI-6) and Incontinence Impact Questionnaire (IIQ-7) were published in 1995, validated, and obtained a grade A recommendation. They are derived from longer questionnaires and both assess the impact of incontinence on QOL. They are the most common clinical tools for women.

Developed by Raz and Erickson in 1992, the SEAPI-QMM (each letter representing an aspect of incontinence) has been grade A recommended, and has widespread use of its questions.

The ICIQ is the most detailed and scientifically accurate tool. It has modules to fit each patient group and is an excellent research tool; however, the applicability for patient use in filling it out is questionable.

The Trial of Mid-Urethral Slings (TOMUS) and the Stress Incontinence Surgical Treatment Efficacy Trial (SISTEr) are highly robust trials. TOMUS utilized the Medical, Epidemiologic and Social Aspects of Aging (MESA) questionnaire, which has a grade C recommendation, and SISTEr used UDI-6, IIQ-7, and MESA.

New symptom scores are being developed, including NNES-Q (Nocturia, Nocturnal Enuresis and Sleep-interruption Questionnaire), which examines nocturia, CLSS (Core Lower Urinary Tract Symptom score), which examines the core of the LUTS score, and VPSS (visual prostate symptom score).

Dr. Badlani questioned if the quest for an ideal questionnaire should be continued. Most questionnaires lack QOL and bother assessments and overlook the patient perspective. He reiterated that observing solely the total score is insufficient.

**Is the AUA Symptom Index Still the Best Instrument in Clinical Research?**
*Claus Roehrborn, M.D., Professor and Chairman, Department of Urology, University of Texas Southwestern Medical Center at Dallas*

In the previous 20 years, nine drugs have been approved for BPH, with the primary efficacy parameter being the AUASS. Many devices and surgical inventions have been approved, but many drugs are ineffective. There currently are 300 active BPH drug/instrument trials. To ensure the effectiveness of an instrument, it should meet quality criteria (measures the properties of health status questionnaires) and be valid.

Research suggests that the IPSS AUA symptom index (AUASI) scale diagnoses BPH, but the overlap is extensive, and therefore, was never intended to be used as a diagnostic tool. In the

original trials, retest reliability was good; however, this since has been disproven in subsequent trials. The AUASI is sensitive to changes and its use is increasing based on the number of PubMed searches (www.gopubmed.org). There are several criticisms of AUASI, however, including that it is not diagnostic, it is subjective, the nocturia question does not correlate with actual nighttime urination, it is unclear what a change in AUASI means, there is a voiding question imbalance, and storage and urge incontinency are not addressed. The AUASI should not be discarded, but it does require some alterations.

**Measuring Incontinence**

*Marcus Drake, M.D., Ph.D., Professor, Department of Urology, Bristol Urological Institute, Southmead Hospital, University of Bristol*

Dr. Drake stated that quantifying leakage underpins decision making for treatment selection and research. For example, a sling or artificial sphincter may be used in mild or severe post prostatectomy incontinence respectively, but there is a failure of consensus on who is more likely to benefit from which procedure as physicians are unable to determine the boundary or overlap between severity grades.

To measure incontinence there are diagnostic tests (mechanistic versus quantitative), leakage tests, diaries, and so forth. Standardized testing to assess severity is difficult because provocation is not easy to standardize (e.g., coughing, straining or exertion), and influences such as fatigue may alter responses. Physiological testing is difficult to standardize due to variations in incontinence (e.g., diurnal, hormonal, post-void residual). Physical variation, differential contribution of other forms of lower urinary tract dysfunction (e.g. detrusor overactivity), and intra-individual variation also render standardization difficult to achieve. Due to these difficulties, researchers and clinicians have not reached agreement on how to best measure incontinence severity.

On the ICIQ for urinary incontinence (ICIQ-UI) short-form questionnaire, one question asks the patient to estimate their leakage. This is a subjective question, which may gain consistency with repeated observation, but intervening consultations also may influence how a patient feels about their urinary incontinence (UI). Likewise, there is a range of patient attitudes, ranging from tolerant ("using four urinary pads daily is okay") to perfection seekers ("any need for pad use is unacceptable"); and these attitudes can be influenced by doctors, websites, and the media. Subjectivity and objectivity create difficulties and numerical quantification would be helpful. For example, pad tests involving measuring the number or weight of urinary pads, can achieve a numerical measure. The practical problems of this are patient embarrassment, behavior adaptation to reduce incontinence severity, and exertion level variability. Leakage detectors exist, but are not as yet reliably accurate at displaying severity of leakage.

The Health Technology Assessment (HTA) programme in the United Kingdom undertook a review of methods of assessing urinary incontinence, which showed a weak correlation between pad tests and scores. Urinary diaries and urodynamic findings are not sufficiently effective (urodynamic findings often do not match patient symptoms). Numerical urodynamic tests for post-prostatectomy stress incontinence, such as the valsalva leak point pressure (VLPP) or maximum urethral closure pressure (MUCP) are not widely accepted as a measure of severity.

Imaging, such as clinical ultrasound, and magnetic resonance imaging (MRI), is not currently a tool for quantifying mechanism or severity of incontinence.

It appears likely that tools will need to be combined for maximum efficacy in diagnosis and treatment. A symptom tool for patient reported bother, a pad test for quantification, a bladder diary to capture situational influence and adaptation, and urodynamic tests for mechanism can be combined to give a general view, but consensus is needed to integrate these into a coherent individualized assessment that can be extrapolated for comparisons. Measuring UI is unreliable and many challenges exist, making the need for a composite evaluation imperative.

**What Domains are Missing in AUASS for Clinical Care?**
*Michael Albo, M.D., Professor, Department of Urology, University of San Diego, California Health System*

The AUA symptom score was developed in 1992 as a tool to quantify the symptoms of bladder obstruction outlet (BOO) caused by benign prostatic enlargement. This score was weighted towards voiding symptoms and was validated only in men. It records the nature, occurrence, and severity of symptoms. Even in the BPH setting, however, the AUASS has limitations, such as a weak correlation with other clinical indicators including nocturia, volume, flow rate, and degree of infravesical obstruction.

The AUASS does not predict outcomes. Doxazosin failed to effectively treat LUTS in 65 percent of men, and while surgery resolved voiding issues, it did not resolve storage symptoms. Thirteen years following TURP, 66 percent of patients had recurrent symptoms, associated mostly with detrusor dysfunction.

A changing concept of LUTS has made symptom scores increasingly problematic. LUTS has evolved into a term that encompasses urinary symptoms regardless of pathophysiology. It is no longer gender specific and comorbidities exist. Classification of symptoms into storage, voiding, and post-micturition can be helpful in focusing evaluation, but cannot be used to diagnose pathophysiology. Thirty percent of patients exhibit multiple symptoms and severity is not necessarily related to bother.

AUASS can be used as a predictor of disease progression as the longitudinal progression of LUTS has been demonstrated in men and women. Symptom change, deterioration, and improvement can, however, vary by patient regardless of the final score. The impact of comorbidities on UI is relevant to consider (e.g., does etiology of LUTS or response vary with conditions and co-morbidity conditions?).

AUASS has problems due to its inability to incorporate symptoms that are associated with LUT function or are caused by treatment. UI is one of the most bothersome symptoms and it is not evaluated in AUASS, neither are sexual dysfunction, pain, anorectal symptoms, and pelvic/prolapsed symptoms in women. Additionally, there is merely one QOL question. Due to these issues, the AUASS is lacking significantly from the patient's perspective. A simple score does not accurately display discrepancies between clinical measurements and symptoms.

**Bother/Adaptation/Exacerbation of LUT Disorders**
*William Steers, M.D., Paul Mellon Professor and Chair, Department of Urology, University of Virginia*

Dr. Steers stated that the one metric that researchers least understand is the one metric that is the strongest predictor of outcome: bother. Bother means to give trouble to, annoy, pester, or worry. Adaptation means to adjust or acclimate to a condition. Remission refers to a state of absence of disease activity in patients with chronic illness. Bother exists in other conditions, such as menopause, fibromyalgia, IBS, and post-traumatic stress disorder (PTSD).

Bother is the most common reason given for treatment of LUTS and is the most important outcome measure. It can be influenced by a patient's character, and may explain the gap between disease prevalence and treatment seeking behavior.

The IPSS bother question has validity and correlates with severity and improvement in symptoms with treatment. In constructing a bother score, the measurement should be precise and reliable, and include normalization for environment, character, and goals. One bother metric asks, "If you spend the rest of your life with these symptoms, how would you feel?" Other bother metrics include patient perception of bladder conditions and a visual analog scale of the bother symptom.

There are age, gender, and racial differences in bother. Storage symptoms create more bother in women than in men. There is less bother associated with each increase in IPSS severity for black men than white. Black men in Africa have less bother than their Detroit, Michigan counterparts for the same LUTS. Predictably, LUTS increases with age, as well as bother; however, bother plateaus at age 60 when adaptation is achieved.

Bother can be altered by other medical conditions. Suffering solely from LUTS creates less bother in men than when accompanied by erectile dysfunction (ED). There is a correlation between depression/anxiety and bother, although it is unclear whether depression/anxiety drive bother or vice versa.

Perceptions are related to bother, such as causal, relative, and uncertainty. The less a patient understands etiology and disease natural history, the greater the bother. This implies that caregivers can influence bother. Bother has links to the disease biology and the biology of bother, caregiver influence, and the natural history of bother.

**Can a Symptom Measurement Tool Tell Us About the Various Contributing Factors to LUTD?**
*John Wei, M.D., Professor of Urology, Department of Urology, University of Michigan*

Dr. Wei explained that symptom measurement tools are not diagnostic; however, clinically, this is highly desired. There is often a mixed etiology and the clinician has to unravel this from a questionnaire.

QOL begins with the biology of the disease leading to symptoms, dysfunction, impairment, and then QOL is affected. Based on this, it should be theoretically possible to decipher the biology of what is occurring based on the symptoms. One way to illustrate this is via an item response theory (IRT). Each item in an item bank describes discrete information about the measurement of any given concept. PRO is an amalgam of these items.

By examining a category response curve or an item characteristic curve from IRT, it can be observed that not all questions are created equal. A good strategy is to determine what is necessary to understand and then develop an item bank. A new questionnaire can be developed based on the contributing factors of interest.

The AUASS exists and typical BOO from BPH is represented but can overlook underlying problems, such as a patient with overactive bladder (OAB). AUASS is a simple and useful measure but is fallible. Dr. Wei displayed several AUASS evaluations that functioned inappropriately for incontinence symptom use.

For each patient, every instrument cannot be used all of the time due to impracticality, and another problem is the usage of condition-specific measures. A good assessment requires administration of multiple measures and the LUTS community has the potential to do better. Existing measures and item banks should move beyond classical test theory and be comprehensive in measurement. IRT should be applied to develop new measures and standardization should be achieved at the professional level.

**How is Post-Prostatectomy Incontinence Measured and What is Missing?**
*Jerry Blavias, M.D., New York, NY*

Dr. Blavias explained that post-prostatectomy incontinence can be defined by underlying symptoms (e.g., stress, urge) and conditions (e.g., sphincter malfunction, detrusor overactivity, low bladder compliance). In addition, the incontinence domain should include its effect on patients (e.g., bother, medical consequences, health-related QOL).

To properly assess incontinence, the following data should be accrued 1) the number of incontinence episodes, 2) estimation of urine loss volume, 3) bother/QOL, and 4) patient assessment of improvement. A bladder diary and questionnaires are positive methods to enumerate incontinence episodes. Volume of urinary loss can be measured by urinary pad counts and weight, as well as bladder diaries, and questionnaires.

A caveat to the above-mentioned methodologies is that the most important tool is the accurate assessment of symptoms, and the currently-used methods may not be measuring symptoms as such. Most of the current methods do not directly measure incontinence. The best outcome tools currently available were developed for use in clinical practices, and the best data entry method is for patients to enter data themselves. Contrary to what some researchers and clinicians believe, this is not too burdensome for patients.

There are three questionnaires that are currently recommended for incontinence, the ICIQ, ICS-male, and DAN-PSS, however, each has major inadequacies and none really fulfills the requirements cited above. The ICIQ modular questionnaires are a well-intentioned idea; however, they do not distinguish between different types of incontinence. The ICS-male examines appropriate symptoms, but the responses are difficult to quantify. A problem with the DAN-PSS is that the distinction between "rarely" and "daily" leaves a large time gap.

Dr. Blavias stated that questions with responses such as "never," "rarely," "few times a month," "few times a week," and "at least once a day" ease patient understanding and offer a reasonable time spectrum. Good questions also include why patients urinate (convenience or severe urgency), how long can a patient "hold it" when they feel the urge to urinate, and the effectiveness of their bladder control.

A bladder diary is a positive method for measuring post-prostatectomy incontinence because it yields many details that a questionnaire does not observe. The longer a bladder diary is maintained and the more systematically they are maintained, the greater the reliability, but the worse the patient compliance. A 24 hour diary is a good compromise. Similar to bladder diaries, pad tests are useful if they are continually done; however, pad counts have not been validated and achieve poor correlation with the volume of urinary loss.

Dr. Blavias concluded that the ideal outcome instrument for incontinence symptoms include a measure of incontinence episodes (bladder diary), volume of urine loss (pad test), patient assessment of improvement (questionnaire), and for stress urinary incontinence (SUI), an examination with a full patient bladder.

***Discussion***

Instead of creating a wholly new symptom measurement paradigm, one aim is to identify ways to move forward with what currently is available. Using a bank of questions is a reasonable option because good questions already have been developed and require minimal modification.

The FDA objective is a request for applications (RFA) announcement for the development of new schemes and ideas to both measure symptoms and phenotype patients. The best ideas would be evaluated in a separate study. This will project quickly forward the necessary accomplishments. Consideration should be given to whether it is best to sort symptoms into bins so that better treatments can be found or to measure the same things as is currently done, but better refine the same measurements. Dr. Badlani commented that such an effort is ongoing and should be improved upon, instead of developing a new questionnaire. It is of note that everything that matters may not be measurable and everything that is measured may not matter; this should be considered in any new or improved methodology.

It is key to develop or improve methodologies so that clinicians can comprehensively and with ease understand and treat patients, even those that have complexities, such as comorbidities. Accumulating knowledge on symptoms, treatments, and outcomes and inserting this data into a database can greatly assist in future research and patient-treatment endeavors.

Dr. Tamara Bavendam, M.D. (Pfizer, Inc.) stated the Pfizer, Inc. developed a comprehensive instrument several years ago during the development of a new drug for male LUTS. Because the drug did not move forward, the questionnaire was shelved. This questionnaire has since been re-examined in the context of overactive bladder treatment and it attempts to give a holistic view of patient symptoms to guide clinicians.

## IV. How are PROs Measured in Other Conditions?

*Moderator: John Wei, M.D., Professor Urology, Department of Urology, University of Michigan*

Dr. Wei indicated that experts from different disease areas will speak so that the field of urinary symptom measurement does not repeat mistakes of other fields or unnecessarily "reinvent the wheel."

### How Do We Measure Symptoms and Flare in SLE?

*Rosalind Ramsey-Goldman, M.D., Dr.P.H., Professor of Medicine, Department of Medicine/Rheumatology, Northwestern University*

The last time that medications were approved for systemic lupus erythematosus (SLE or "lupus") was in 1955. Dr. Ramsey-Goldman's work focuses on studying lupus damage, disease flares, and assessing when conditions are worsening.

Even though lupus is taught in medical school, many people are unfamiliar with it. It is a chronic, multi-system, autoimmune, inflammatory disease. It is characterized by exacerbations (flares) and remissions and immune dysregulation (e.g., loss of tolerance to self-antigens, production of auto-antibodies). Nearly all lupus patients are women of child-bearing age; however, the ratio of male to female patients is nearly equal at the age extremes. Symptoms of lupus include a butterfly rash on the face, discoid rash, cytopenias, pleuritic chest pain, arthritis, and leg swelling which can reflect kidney lesions. These problems can result in organ damage such as end-stage renal disease.

Lupus is measured by a disease activity measures which comprise both clinical and laboratory assessments of a physician's global assessment (PGA) which is an overall estimate of disease activity. In addition, damage is assessed using the systemic lupus damage index (SDIAn activity assessment (systemic lupus erythematosus disease activity index [SLEDAI]) is often used. Other instruments that can be used are systemic lupus activity measure (SLAM), systemic lupus activity questionnaire (SLAQ), and British Isles Lupus Assessment Group (BILAG). SLEDAI functions by scoring based on the presence or absence of weighted variables. Although SLEDAI is simple, it has the disadvantage of not accounting for change in severity or accounting for rare disease manifestations. Some symptoms worsen while others improve, and the global SLEDAI score does not make this apparent. The BILAG instrument is scored using the principal of intention to treat the disease. Advantages include organ specificity, assessment of severity, and accountability of change. Disadvantages include a complicated scoring system and a threshold effect across different organ systems. PGA supplies an overall physical impression with the

advantage of being easy and simple to administer. A disadvantage is that is overlooks individual organ system changes. Determining which one is optimal requires researchers and clinicians to find a consensus; however, all have undergone revision and modifications for use in clinical trials. Common to all indices are psychometric testing, definition of terms, sensitivity to change over time, and training to use instruments. Biomarkers and PROs also can be used to measure disease activity, prognosis or response to treatment, but these are tools which are currently either in development or being tested for use in lupus.

There are many challenges to using disease activity to quantify flares in lupus, including inter-patient disease heterogeneity, unpredictable intra-patient waxing and waning of the disease, discordance between clinical symptoms and serologic tests, and limitations of disease activity measures that were developed for research being used to assess response in clinical trials. Disease activity and flare are different because activity encompasses all symptoms and signs related to SLE pathophysiology, while flare is defined as an increase in disease activity as compared to previous assessments.

Both the BILAG (version 2004) and SLEDAI have been adapted for flare assessment in flare studies. In examining if these instruments reliably identify flares and flare levels when used in a lupus clinic, a London study found that 84 percent of the time the two measurements agreed. Fifty percent of the time they agreed on the type and level of flare with the best agreement for severe flare. However, more work is needed to capture clearly a mild versus moderate flare.

**What Tools are Available for Symptom Measurement in IBS?**
*Lin Chang, M.D., Professor of Medicine, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles*

Irritable bowel syndrome (IBS) has been defined as "a functional bowel disorder characterized by abdominal pain or discomfort that is associated with altered bowel habits, that is diarrhea and/or constipation," Dr. Chang indicated that a key clinical finding in IBS is a link between pain and bowel habits.

The Rome III symptom-based criteria, developed by the Rome Foundation, are used to diagnose IBS. These criteria require the presence of abdominal pain three or more times a month for at least 3 months, that is associated with at least two of the three features: improved with defecation, associated with a change in stool frequency, and/or associated with a change in stool form. IBS patients experience multiple symptoms including nausea, bloating, gas, fullness, and urgency. The most bothersome symptom can vary among patients. Most treatments target the normalization of bowel habits and reduction of pain.

The Bristol Stool Form Scale is a validated measure that is used to subtype IBS patients by predominant bowel habit, (e.g., IBS with diarrhea, constipation, or a mixed pattern). This form is used in both clinical practice and research.

Objective markers for IBS are being studied, although currently there is no biomarker that has been shown to reliably and consistently diagnose IBS. Potential biologic markers which have

been studied in IBS research studies include serum biologic markers, visceral sensitivity using barostat measurements, gastrointestinal transit times, immune responses, brain imaging patterns, and autonomic nervous system responses. PROs are currently being used to measure symptoms in IBS. IBS severity has been measured using different PROs including the IBS symptom severity scale (IBS-SSS). IBS severity is a multi-factorial concept and accounts for many issues (e.g., gastrointestinal symptom intensity, psychosocial factors, impact on life, stigma, coping).

Efforts are being made to develop a valid PRO for IBS using the FDA guidelines for PRO development, such as those by the IBS working group which is part of the C-Path PRO Consortium. There is also a NIH funded PROMIS project that is developing a GI symptom scale which can be used to assess gastrointestinal symptoms across conditions and in the general population. In the meantime, the FDA has recommended the use of interim composite endpoints for IBS clinical trials.

**Discussion**

Dr. Chang responded to a question regarding the connection between the diagnosis of IBS subtypes and outcome measures by stating that IBS diagnosis is based on the presence of abdominal pain with diarrhea or constipation. Therefore, a PRO evaluates both abdominal pain and bowel habits but could conceivably be different in IBS with constipation and IBS with diarrhea.

For IBS, researchers and clinicians generally design instruments based on GI symptoms. Dr. Chang clarified that IBS is a chronic or recurrent pain syndrome, and therefore pain can be assessed by other chronic pain PROs although they would have to be validated in the IBS population. In the past, pain and discomfort have been grouped together both in the diagnostic criteria and in PRO questions which should change since many IBS patients perceive pain and discomfort as two separate entities.

**Symptom Measurement in FM/CFS**

*Fred Friedberg, Ph.D., Research Associate Professor, Department of Psychiatry, Stony Brook University*

Fibromyalgia (FM) is a chronic pain condition that is characterized by unexplained widespread pain that lasts longer than 3 months. This often includes exquisite pain sensations produced from gentle touches. The 1990 criteria used 18 tender points and if a patient felt intense pain in 11 of those areas, they were diagnosed. The new FM criteria do not require tender points.

Chronic fatigue syndrome (CFS) is characterized by 6 months of persistent fatigue, substantial impairments (e.g., occupational, family responsibilities) and several secondary symptoms, including flu-like symptoms, pain symptoms, unrefreshing sleep, and post-exertional malaise. The post-exertional malaise is a unique symptom of CFS. The fatigue is not resolved by sleep or exercise. FM and CFS have a comorbidity rate of 50 percent. Patients usually appear healthy and the causes are poorly understood.

Symptom assessment is critical because there is no examination to determine if a patient has the illness. Symptom measures for FM/CFS should assess the symptom intensity/impact, determine symptom location, as well as assist in the delineation of illness subcategories (high vs. low functioning).

Single-item measures of fatigue and pain are often used and include numerical ratings, a verbal rating scale, and a visual analog scale. Pain and fatigue can be monitored over time which can assist in determining the effectiveness of interventions. The comprehensive McGill Pain Questionnaire (MPQ) measures three dimensions of pain (sensory, affective, and evaluative). The Multi-dimensional Pain Inventory (MPI) examines pain intensity, emotional distress, cognitive and functional adaptation, and social support. Factor analyses have revealed three FM pain profiles (i.e., dysfunctional, interpersonally distressed, and adaptive copers). The brief pain inventory (BPI), FM assessment status index (FAS), FM impact questionnaire (FIQ), and the revised FIQ are commonly used assessments. The revised FIQ has been well-validated and can discriminate between lupus, depression, and FM.

For CFS, the brief Fatigue Scale (FS) is most commonly used and is well-validated, although it has a possible ceiling effect. Other assessments include the Multi-dimensional Assessment of Fatigue (MAF) and the Multi-dimensional Fatigue Inventory (MFI-20). Self-report measures are problematic because there are varying levels of psychometric validation and not one scale has been accepted as the standard. The PROMIS symptom measure is relevant to symptoms assessment in FM and CFS; however, it does not adequately measure post-exertional malaise.

**UCPPS, Measuring Pain, and Flares**

*J. Quentin Clemens, M.D., Associate Professor of Urology, University of Michigan Medical Center*

Urological chronic pelvic pain syndromes (UCPPS) is an informal term for IC/bladder pain syndrome (IC/BPS), as well as chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS). Pelvic pain occurs in both males and females and includes bladder and non-bladder pain. CP/CPPS is pain in the absence of a UTI or other identified cause, and is exacerbated by urination or ejaculation. IC/BPS is characterized by an unpleasant sensation (e.g., pain, pressure, and discomfort) perceived to be related to the urinary bladder and is usually associated with LUTS.

The diagnosis of CP/CPPS is typically pretty straightforward as other causes of male pelvic pain are uncommon. IC/BPS, conversely, is controversial because it can sometimes be difficult to distinguish from other disorders (e.g., overactive bladder, endometriosis, vulvodynia). Furthermore, IC/BPS can coexist with other pain conditions such as endometriosis, vulvodynia, etc.

Validated instruments include the NIH Chronic Prostatitis Symptom Index (with pain, urinary, and QOL subscales), the IC Symptoms and Problem Index, and the NIH Genitourinary Pain Index (GUPI). These questionnaires measure symptoms severity and are not designed to be

diagnostic. Conversely, there are diagnostic questionnaires have been developed to identify the presence of IC/BPS or CP/CPPS with a known sensitivity and specificity.

The Multi-Disciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) network is a multi-site study funded by the NIDDK which is focused on identifying relevant disease phenotypes in men and women with UCPPS. One particular focus if the MAPP project is to examine symptom changes over time and to better understand the characteristics of disease 'flares'. Preliminary data suggest that there is significant variability across patients in these characteristics, and future studies may help to identify factors that predispose to symptom flares. Additional research questions how symptom variability affects QOL, the exact meaning of "flare," and what patients would describe as a flare. Focus groups also are being formed to identify missing domains.

Concerning conceptual groups, sensory/afferent abnormalities can be distinguished from motor/efferent abnormalities and structural abnormalities. Objective signs and tests are elusive for afferent symptoms. The overlap of these three conceptual groups creates difficulty in assessment development.

***Discussion***

A participant queried the development of disease concepts on which to structure questions. Dr. Friedberg commented that FM/CFS comorbidities and patient diagnoses stemmed this development; few tools are specific to CFS and FM. Dr. Ramsey-Goldman indicated that her challenge was discussing flare, which only involves physical measures, even though many instruments are used from patient symptoms. She continued that a definition of flare was obtained via iterative questionnaires, with the goal being to have a questionnaire usable in practice.

When a patient struggles with IC, deciding whether to have LUTS or LUTD examined is best determined by talking with the patient. This practice acknowledges what is important to the patient, and if communication issues can be bypassed, then a comprehensive understanding of LUTS or LUTD can be determined.

A participant questioned the assessment of lupus flares from the disease state, as well as observing disease activity and attempting to induce remission. Dr. Ramsey-Goldman commented that a responder index was needed. SLEDAI measures global disease, but it is important to ensure that therapy does not worsen a different condition. A combination of a responder index, SLEDAI, and BILAG allows for an overall assessment of patients. Clinical trials use this strategy.

## V. PHENOTYPES, CLINICAL CONDITIONS, AND IMPACT ON MEASUREMENT

*Moderator: Michael Albo, M.D., Professor, Department of Urology, Bristol Urological Institute, Southmead Hospital*

### Healthy Aging. The Impact of LUTD.

*Tomas Griebling, M.D., M.P.H., Wolf Masonic Distinguished Professor of Urology, Department of Urology, University of Kansas*

Dr. Griebling discussed how incontinence influences aging and geriatric issues in older adults. Issues to consider include epidemiology, QOL, frailty, geriatric syndromes, and so forth. In the United States, the social definition of geriatric generally applies to people over 65 years of age, although the WHO defines geriatrics as individuals over the age of 60. However, new definitions are elevating the age of geriatrics to more than 75 years of age because people live longer and healthier lives.

Addressing gerontology and incontinence is a demographic imperative; the elderly comprise 13 percent of the U.S. population and this number could rise to 20 percent over the next 20 years. Over the past 40 years, the number of people in their older years has steadily increased with time. This phenomenon is not limited to the United States; countries within Europe, North America, South America, and Asia also contain increasing geriatric populations. Only in sub-Saharan Africa is this not the case.

The prevalence of incontinence increases with age, but it should not be considered an inevitable or normal part of the process. Types of UI that occur in older adults include urgency, stress, mixed, and other (e.g., fecal). Regarding QOL, UI is considered a marker condition for morbidity or mortality, is associated with a two-fold increase in depression, negatively impacts sexual health, and results in impaired self-reported health.

Incontinence costs in the elderly also affect personal caregivers (i.e., direct and indirect costs, containment products [pads], and environmental costs).

A question exists whether UI is a clinical diagnosis or a geriatric syndrome. Of elderly people with UI, 60 percent had at least one other geriatric syndrome. Risk factors for UI includes nocturia and fall risk—nocturia increases the risk of falls by 10 to 20 percent in older adults. The fall risk is multifactoral and includes urgency and the physical limitations of navigating to a toilet. There is an increased risk of hip and long bone fractures, and increased risk of both morbidity and mortality. The risk is higher in both men and women with osteoporosis or osteopenia.

Dementia, or cognitive impairment, has been associated with “functional” incontinence. There is an increase in detrusor overactivity, although that may be due to normal aging. Cholinesterase inhibitors may worsen incontinence making the role of anticholinergic medications controversial.

UI is not necessarily predictive of nursing home placement; study results have been mixed. However, UI is prevalent in nursing home patients (30 to 50%) and the risk of new onset UI is fairly common. Treatment of UI in the form of assisted toileting can be effective.

Frailty is the sum of changes associated with increased impairment in older adults. Frailty is indicated by decreased walk time, slower gait speed, diminished grip strength, decreased physical activity, easy exhaustion, and unintended weight loss. The impact of incontinence on longevity, functional reserve capacity, survival healing, and other factors are under investigation.

UI is often accompanied by comorbidities. Additionally, UI is associated with self-reporting of poor health. In a national survey of self-care and aging that measured activities of daily living, morbidity, and mobility, patients who were incontinent possessed a three-fold increase in the odds of reporting poor health. Regarding the relationship between UI and death, there is an increase in moderate and severe risk.

The Vulnerable Elders Survey (VES-13) does not specifically address incontinence, but examines other functional components. MESA and the Sandvik severity index are other measures of incontinence. Objective measurements include pad testing, urodynamics, history, and physical examination. Functional assessments address function (e.g., independence, disability), and activities of daily living (e.g., dressing, bathing, grooming, shopping, housekeeping, accounting, food preparation, telephone communication, transportation, and taking medications). These evaluations are important because average life expectancy and annual health care costs are correlated.

Mobility is assessed with the Get Up and Go Test, which measures gait and balance. The patient sits in a straight-backed chair, stands, walks 10 feet, turns, returns to the chair, and sits back down. This measurement is untimed and personal assistive devices are allowed.

A cognitive assessment is helpful in understanding preoperative baseline function. Delirium is evaluated after surgery. The Folstein Mini Mental State Exam (MMSE) is well validated, although many clinicians now use Mini-Cog, which is a three-item recall and clock-drawing exercise that tests short-term memory and executive-level function.

Dr. Griebling emphasized that although incontinence is common in both men and women, unique gender differences are important. Incontinence is considered a specific diagnosis, but in the context of a geriatric individual, it can have an impact on general health as well.

**What Did We Learn From MTOPS?**

*Kevin McVary, M.D., Professor, Department of Urology, Feinberg School of Medicine, Northwestern University*

Dr. McVary introduced MTOPS as a successful urology story. He described the natural history of BPH as an increase in prostate volume over time as individuals age. In some people, the prostate volume changes quite minimally, while the volume increase is quick and dramatic in others. This raises a fundamental phenotype question: Do clinicians treat both cases the same?

In predicting which patients will progress, Dr. McVary noted several questions that need to be answered. Who will experience the accepted level of change? Who will worsen? Who is at risk for retention? Who will need invasive intervention? These questions were addressed with the MTOPS clinical trial.

Prior to MTOPS, it was known that taking alpha blockers make a five to 10 point difference in a year. Finasteride provided a three to four point difference in 4 years and was well-tolerated. Some long-term effects on progression were known, but there was no information regarding long-term use of alpha blockers. The aim of MTOPS was to determine the long-term history of progression specific to the application of a combination of therapies and to determine if disease progression was inhibited or worsened.

MTOPS screened 4,000 patients and randomized 3,000 of them into different treatment arms. The primary endpoint was a composite measure detecting the first occurrence of any of the following five outcomes: (1) four-point rise from baseline in AUASI; (2) acute urinary retention (AUR); (3) renal insufficiency due to BPH; (4) recurrent UTI; and (5) urinary incontinence. Secondary outcomes included crossover to invasive therapy and other biological measures. There was a high level of adherence and compliance over the 5-year trial; only 11 percent discontinued the trial.

In the placebo group, patients experienced BPH progression (17%), a four-point AAUSS rise (14%), AUR (2.4%), and UTI. The secondary endpoint of BPH invasive therapy affected 5 percent of the patients. The cumulative incidence of BPH progression indicated that monotherapies led to less BPH progression than placebo, with the combination therapy being superior to both monotherapies.

A U.S. Department of Veterans Affairs (VA) Cooperative study found different results, but the VA study concluded at 1 year while MTOPS studied the long-term perspective. Indeed, the MTOPS data concur with the VA data in the first year, but the data diverge when followed longer.

AUR is as important and bothersome to men as is a stroke or myocardial infarction. Alpha blockers are not a long-term therapy to prevent this outcome. Finasteride alone or in combination is superior to prevent this endpoint. Regarding the men who crossed over to BPH invasive therapy, alpha blockers do delay the crossover, but at the end of 5 years, the result was the same as for the placebo group. Based on the PSA values that were used as a proxy for prostate volume, combination therapy is the best at preventing clinical progression.

AUR was stratified by prostate volume. For large prostates, placebo and doxazosin had similar effects. Finasteride alone and together with doxazosin was superior. The same effect was documented for intermediate-sized prostates. Thus, for large prostates combination therapy is indicated, and for medium-sized prostates combination therapy may be preferred. However, with small prostates, all of the medications impact progression so there is no need for combination therapy.

MTOPS was useful in determining the proper form of combination therapy. It was discovered that alpha blockers do not delay everything, but 5-alpha reductase inhibitors do have the power to influence disease. Managing LUTS is important to limit progression and the effect of the disease over time.

In terms of answering the questions, MTOPS did not clarify which patients will experience accelerated prostate volume, but made some progress toward identifying who will develop worse symptoms. Age, PSA levels, and total prostate volume (TPV) are important factors influencing progression. These factors, along with the baseline infiltrate on transrectal ultrasound prostate biopsy (TRUS-bx), were also shown to have influence. Finally, the question about who needs surgical intervention was not addressed because that is a patient choice.

**Can Symptoms be Predictive of Progression and Response to Therapy?**
*Aruna Sarma, Ph.D., Associate Professor, Department of Urology &Epidemiology, University of Michigan*

Dr. Sarma considered if clinicians can predict progression and treatment in LUTS, and suggested an affirmative answer. Population-based studies have found that LUTS increase with age, although the MTOPS study indicated that the rate of progression is not as steep as once thought.

The AUASI tracks some changes over time, but is the question whether assessment drives symptoms, or the symptoms drive the assessment? There is some amount of variation from these studies; moderate to severe LUTS can progress to severe, but some regress, so patients move across categories and not in one single direction. A publication several years ago indicated that people with LUTS go in many directions.

Long-term changes are calculated by dividing the baseline symptoms score by the time and bother score. There are concomitant changes that accompany bother; generally, changes in bother are correlated with changes in the progression of LUTS. Outliers exist, such as stoic patients and those who have difficulty accepting the symptoms. Predicting those individuals is difficult.

In graphs that chart the odds of having a symptom, the most important values to investigate are those at the extremes, which have significantly greater odds of outcomes. Symptoms predict AUR. Individuals who have a symptom index of eight or higher have more than three-fold odds of getting AUR. Symptoms also predict treatment. After adjusting for age and other variables, there is a five-fold increase for patients experiencing symptoms to report a treatment.

An important question to understand is what is being measured. Not all LUTS are equivalent and other symptoms must be taken into consideration. The possibility of a diagnosis of a similar but distinct disease must be acknowledged. Medications such as antihistamines and anticholinergics also may present symptoms similar to LUTS.

LUTS can be used to predict progression and treatment, but variability exists. Some variability is due to the underlying etiology, but other variability is due to an unrelated genesis.

**Symptomatic LUTD, ED, and Metabolic Syndrome: What Is the Relevance?**
*Aruna Sarma, Ph.D., Associate Professor, Department of Urology &Epidemiology, University of Michigan*

Dr. Sarma discussed the importance of understanding the relevance of the relationship between LUTS, ED, and metabolic syndrome (MS). It is known that LUTS and ED are both prevalent, increase with age, and decrease QOL. However, a causal link has not been identified yet, and the role of MS is unknown. MS is reaching epidemic proportions, with almost half of all people experiencing some component of MS. This underscores the need to understand the syndrome more completely.

MS is the name for a group of risk factors that contribute to heart disease, diabetes, stroke, and other problems, and is primarily diagnosed by ATP3 levels. MS has manifestations involving insulin resistance and varying degrees of MS have been associated with BPH/LUTS.

ED increases with age and the proportion of patients with ED increases with LUTS severity. The Olmsted County study found similar correlations. Natric Oxide Synthesize (NOS/NO) levels have been documented to be decreased or altered in the prostate and penile smooth muscle. Pelvic atherosclerosis may be a mechanism for LUTS and ED. Autonomic hyperactivity and metabolic syndrome affects LUTS, and prostate growth on ED. Components or outcomes of MS can be associated with ED and LUTS.

LUTS are a primary manifestation of BPH, but not only caused by BPH. The AUASI is used to identify BPH and LUTS, but the limitations on these assessments are that they are highly sensitive but not highly specific.

Current data indicate the urological complications of diabetes, which pose a burden to patients and the health care system. The Diabetes Control and Complications Trial (DCCT) / Epidemiology of Diabetes Interventions and Complications (EDIC) Study, performed in the late 1980s, found a large reduction in issues in response to intensive glycemic therapy. In 2003, NIDDK assessed urological symptoms. They identified patterns of LUTS with men as they age, but those with glycemic therapy had more LUTS, not less. Current analyses of the data indicate that obesity may explain part of the result. Metabolic syndrome by itself is just a syndrome, but from a population perspective, a patient probably does not just have diabetes, but they are diabetic, obese, and so forth. It will be important to discern how to address patients with multiple symptoms.

Not only do physical complications change, but people who have ED and LUTS report a decreased QOL. This provides a good reason for why these symptoms should be studied in comorbidity patients—they impact QOL. These urologic factors are a reason for people to lose weight; although they are not as serious as diabetes-related problems later in life, they are important earlier. Additionally, ED is related to LUTS in men with diabetes.

Phosphodiesterase type 5 (PDF5) inhibitors impact ED and LUTS. There is a relationship between LUTS and ED, with biologically plausible links and strong associations, such that

treatment of one impacts the other. There are an increasing number of affected men in the aging population, and many LUTS treatments (e.g., medications and surgery) affect sexual function. Conversely, ED therapies may improve LUTS.

The global challenge is to understand LUTS and ED, and an emphasis should be placed on a systems biology approach that develops a common and integrated measurement and management approach for LUTS and ED, keeping metabolic syndrome in mind.

**DM and LUTD Symptoms**

*Firouz Daneshgari, M.D., Lester Persky Professor and Chair, Department of Urology, University Hospitals*

Dr. Daneshgari discussed the relationship between diabetes mellitus (DM) and LUTD symptoms, elaborating on three main points: (1) DM is contributing to an increasing prevalence of LUTS in men and women; (2) the temporal effects of DM affects the full spectrum of LUT function, including storage, voiding, and post-voiding; and (3) diabetic bladder dysfunction represents a two-way transitional model for studies of LUTD in that it affects both genders, there exist loyal animal models, the diseases/symptoms are responsive to prevention and treatment, and a community-based approach can be applied.

The estimated lifetime risk of an individual to develop diabetes is very high, a diagnosis that includes a significant cost and burden. The age at which a person is affected by diabetes is decreasing. According to data from the Centers of Disease Control and Prevention (CDC), most new cases occur in middle-aged individuals (40 to 60 years old), and the biggest proportion of the new cases are Type 2 Diabetes (T2D), as opposed to Type 1 Diabetes (T1D).

The prevalence of urinary incontinence was studied in a 7-year period. Close scrutiny of the data revealed that after adjusting for women and diabetes, the difference is almost entirely contributed by the role of diabetes. Urological complications of obesity and diabetes (UCOD) include bladder dysfunction (DBD), UTI, and prostate enlargement. UCOD is the largest contribution to complications.

The current state of knowledge can be represented by two cases. The first case involves a 52-year-old woman with T1D who has UI (i.e., overflow incontinence). Due to the effect of diabetes on the bladder, patients have an enlarged bladder and decreased sensation. The second case involves a similar-aged woman with T2D. Her urinary incontinence was due to urodynamic detrusor overactivity. Bladder overactivity is a prevalent phenotype in diabetics; especially in T2D, if clinicians look for it.

Based on data from experimental studies, Dr. Daneshgari proposed a natural history study of LUTD in DM patients. The study would go through two phases, one being early presentation in the clinical level by storage bladder problems, evidenced by urodynamic detrusor overactivity, and the other being in the late phase with voiding problems, and atonic bladder. It is known in animals that the transition timetable from a compensated or storage problems to decompensated

or voiding problems is between 9 to 12 weeks, but it is important to identify the appropriate timetable in human.

It is difficult to determine clinical understanding of basic science discoveries by examining the natural history of diabetes because when patients develop symptoms of LUTS, they migrate from endocrinologist to urologist doctors.

The important research questions include identifying the time lag between the early and late phases, finding differences between T1D and T2D, and discovering if there is a "point of no return" past which the individual is fated to an end staged bladder.

Clinical issues resulting from DM include that the bladder undergoes marked remodeling, the net effect of modeling is hypertrophy, there is altered afferent bladder sensation, urinary incontinence, detrusor overactivity, and failure of the bladder to empty. The reality is that a full-tool spectrum exists from the phenome to the genome. Examination of tissues, metabolites, proteins, and transcript levels allow researchers to go from the phenotype in the bladder (remodeled) and hypothesize mechanisms (e.g., the early phase is driven by polyuria and the late phase is driven by hyperglycemia).

The prevalence and progression of urinary incontinence also has been studied in women. Women with DM have elevated prevalence of incontinence. DM causes more severe stress urinary incontinence (SUI) in women. With this observation, the association between pregnancy induced birth trauma and SUI warrens investigation. The proposed study would investigate the prevalence and progression of prenatally-associated SUI among diabetics and obese women.

**How to Better Phenotype Patients With Symptomatic LUTD**
*William Steers, M.D., Paul Mellon Professor and Chair, Department of Urology, University of Virginia*

Dr. Steers noted that through phenotyping patients, great sensitivity and poor specificity of outcome measures could be clarified.

LUTD patients are heterogeneous making the grouping of them problematic. Patients should be parsed instead of clustered.

Through discussions, it was concluded that the symptom complex in a heterogeneous population may comprise the same symptoms. When discussing phenotyping, one definition combines *in vivo* evaluations and pathology to observe complex phenotypes, which are validated with genetic traits. It is important to venture into the field of epigenomics. The genotype is influenced by environmental and natural interactions in the development of a phenotype in patients. Stress and injury impact LUTD, indicating that it is not a simple disease.

Current therapies are empirically based on symptoms and bother. Treatments are mediocre by targeting the bladder as an "end organ." There is less rationale to phenotype than if the symptoms affected the central nervous system, vascular system, and so forth. Therapies fail to

target the underlying cause or reverse the disease process. Phenotyping may provide insight into pathophysiology and may improve outcomes in some treatments that work dramatically well in a few patients. Other patients do not find improvement indicating that the response shows no efficacy.

In phenotyping IC or painful bladder syndrome, if a cohort is grouped together and treated with an agent, often no efficacy is seen. A LUTS phenotyping study examined sex, history, and so forth. Urinary symptoms documented included nocturia, urgency, frequency, incontinence, pad usage, diary documentation, and Likert scores. Psychosocial and character assessments monitored depression and anxiety, OCD, ADHD, and PTSD. Populations with these traits may respond differently to therapies. Neurological conditions and systemic disease (e.g., dementia, MS, DM, peripheral vascular disease) should either all be excluded or all included in clinical trials.

Dr. Steers stated that symptoms should indicate which patients require treatment, and PSA values indicate the proper treatment methodology.

Childhood disorders (e.g., recurrent UTI, constipation) have not garnered good longitudinal studies or registries at this time.

Predictive phenotypes can indicate success (e.g., duration or severity), failure (e.g., neuropathic, DM), durability, progression (e.g., DM, MS), and remission.

There are early signals in animal models for fibroblast growth factor and adrenoreceptor effects; however, trials have not been designed for humans. Predicting the effects of these on human patients based on the underlying biology would be helpful.

Only by phenotyping can the FDA goal of prescribing the correct medication for the correct patient at the correct dose be achieved. Rather than starting "from scratch," the LUTS field already contains metrics and data from large trials which needs to be located in a central data repository.

***Discussion***

The idea that Finasteride decreases prostate size is considered simplistic by some. Rapidly after the onset of Finasteride use, endothelial cell apoptosis occurs. When the prostate decreases in size, the tensile characteristic of the stroma alters.

Practical phenotyping can occur if a hypothesis is developed initially. Some phenotypic evaluation data that have been found in studies have not been warehoused, and so a full evaluation and information range could be made available to assist in future phenotypic evaluations. Phenotyping is promising, but must be complemented by *improved* phenotyping and taking advantage of clinical database resources.

Regarding diabetic neuropathy, patients with a full bladder, atonic bladder, and so forth had much neuropathy which led to sensory deficits. Sensory deficits are examined via a decreased sensation clinical measure.

In men, when an overactive bladder is associated with obstruction and the obstruction is relieved, OB symptoms improve. It is not advantageous to group clinical phenotypes and to tease apart these issues at the clinical level, separate phenotypes must be observed and then later potentially combined.

It is positive news that the pharmaceutical industry realizes the declining viability of a "blockbuster" model. This will improve the efficacy and exactness of medication use; however, not all afflictions will garner an approved medication because of the high number of sub-categorized afflictions.

## TUESDAY, NOVEMBER 15, 2011

### VI. BETTER UNDERSTANDING OF THE SYMPTOMATIC LUTD PATIENT: THE FUTURE

*Moderator: Kevin Weinfurt, Ph.D., Professor, Department of Psychiatry and Behavioral Science, Duke Clinical Research Institute, Duke University*

### What Measures are Needed and for What Purposes?

*Mathew Barber, M.D., M.H.S., Professor of Surgery, Department of Obstetrics and Gynecology, Section of Urogynecology and Reconstructive Pelvic Surgery, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University*

Many outcome measures have been developed for clinicians, researchers, and clinical trials. The AUASS is a symptom score that is widely used outside of its intended purpose. There are, however, many instruments available to study LUTS. The best of these instruments should be compiled to form a LUTD measurement "toolbox," in which enough tools are present to meet the needs of researchers and clinicians, but there is not an abundance that can lead to confusion.

Surveying men and women with LUTS, there is considerable complexity in symptoms and overlap. The largest patient group includes voiding storage and post-micturition symptoms. There are differences by gender, as well (e.g., males often have voiding function problems, while females often have storage symptoms). Among storage symptoms in men and women, overactive bladder is the most common. Additionally, women have a higher prevalence of UI due to child bearing, while the proportion of men with UI increases with age.

Gender-specific measures are needed due to differences in symptoms, natural history, daily function, interpretation and description of symptoms, and differential interpretation of questionnaires. Before an instrument is used in a population of men and women it should be demonstrated that instruments have same properties/interpretation in both men and women before adoption across both populations.

The Wilson Cleary model is the optimal model of continuum health measures. Clinicians are concerned with symptom status related to biology, while patients are concerned with the opposite end of the spectrum, QOL. When using PROs it is imperative to determine the use of the PRO and the person analyzing it (e.g., clinician, industry, government, clinical researcher, and so forth).

In measuring LUTD, there are multiple etiologies with a wide variety of clinical presentations, which implies the need for a broad spectrum of instruments. An imperfect understanding of the biology of underlying symptoms adds to difficulties (e.g., OAB is a symptom, not a disease).

Diagnostic screening instruments and outcome measures are both needed, but the qualities of a good diagnostic test and a good outcome measure are different so it is rare that one instrument can adequately serve bother purposes. Patient outcome assessments are helpful because many treatment effects are known only to patients (e.g., bother, QOL). Most cases of LUTD do not result in morbidity or mortality; the impact is on QOL, and so this is important to properly assess outcomes.

A useful PRO will include patient perspectives, psychometric properties, be easy to understand and implement, and be interpretable by patients. The mode of administration and language translation also are important to consider.

There is an imperfect relationship between severity and health-related QOL (HRQOL; R=0.26). Only 5 percent of patients with urinary incontinence have bother. Symptom-specific and global measures for LUTD outcome measures are clearly needed.

Clinicians require brief, simple questionnaires with a broad symptom measure, assessment of patient goals, and diagnostic instruments. Researchers require a broad array of PRO assessing LUT diseases specific to populations and outcomes of interest. Many validated instruments representing many domains are available; there is not a need to invent new ones. Advanced methodologies can provide harmonization, and direct scientific comparison of different measures should be used to identify the best instruments.

**Should Symptoms of LUTD in Men be Measured Differently in Studies Addressing Modifiable Risk Factors?**

*Elizabeth Platz, M.P.H., D.Sc., Professor and Martin D. Abeloff, M.D. Scholar in Cancer Prevention, Department of Epidemiology, The Johns Hopkins Bloomberg School of Public Health*

Regarding the identification of modifiable risk factors for the development and progression of BPH/LUTS, the purpose is to intervene and reduce risk. Intervening on risk factors can help prevent or intervene in other diseases as well. For this to occur, BPH/LUTS definitions are required, and they must be measurable.

Traditionally, a pharmacologic approach is used to reduce symptoms, but a new non-pharmacologic approach that involves intervention to prevent incidence is optimal. Dr. Platz

conceded that all LUTS cannot be prevented, but onset could be delayed and progression could be slowed.

An NIDDK-sponsored roundtable discussion in 2008 focused on defining LUTS/BPH in observational epidemiologic studies. This was important to increase comparability between studies and to facilitate the grant application review. Epidemiologic studies use thousands of men, and some measurement tools in trials may not be feasible for large cohorts. Additionally, large cohorts are expensive to develop using labor-intensive measurements. Existing cohorts should be utilized because they have a long-term follow-up and have been measured for LUTS/BPH many times with multiple timepoint collections. This is a rich resource that must be exploited.

The definition of LUTS varies depending on the study scale. Large and small studies use interviews and medical record review. Small studies use urinary flow testing, blood-based biomarkers, and urine-based biomarkers. Case studies use tissue-based biomarkers.

There are several points of agreement, however, including the acknowledgement that BPH/LUTS is a complex set of conditions with many underlying pathologies that can yield that same symptoms, other conditions may masquerade as BPH/LUTS, the currently used definitions and measurements are heterogenous, and both broad and narrow definitions of BPH/LUTS may have utility in uncovering modifiable targets for intervention. The next steps that should be taken include the development of a conceptual model and a global name for the constellation of conditions that comprise BPH/LUTS, and the measurement and observation of phenotypes. The ultimate goal is to reduce the burden of BPH/LUTS at the population and individual levels via intervention throughout the course of a life.

**Getting to the Core of Cancer Outcomes Measurement**
*Claire Snyder, Ph.D., Associate Professor, Department of General Internal Medicine, The Johns Hopkins University School of Medicine*

Cancer research and practice have been a major focus for assessing health-related quality of life (HRQOL) and other patient-reported outcomes (PRO). Identification of core measures and measurement instruments has been long discussed. In 2001, the National Cancer Institute (NCI) formed the Cancer Outcomes Measurement Working Group (COMWG) to assess the state-of-the-science of outcomes assessment in cancer and to make recommendations for moving the field forward. The 35-member multidisciplinary working group addressed measurement of three outcomes (HRQOL, patient needs and satisfaction, and economic burden) in four cancers (breast, prostate, colorectal, and lung) across the cancer care continuum (prevention and screening, treatment, survivorship, and end-of-life), and addressed multiple applications including clinical practice, research, and policy. In the end, the COMWG determined that selecting core measurement instruments was not appropriate because: (1) each study needs measures appropriate to its distinct hypothesis and patient population; (2) it was too early in the field's development to identify core measures; (3) although there were many measurement instruments that met the criteria for a gold standard, no instrument stood above the rest; and (4) the advantages of innovative methods such as item-response theory that allow comparisons across

measurement instruments could make selection of the core unnecessary. The COMWG suggested identification of core concepts to measure rather than core measurement instruments.

Following on the COMWG's work, the Patient-Reported Outcomes Measurement Information System (PROMIS) sought to develop item pools that can be used to create static short forms and adaptive tests for key domains applicable to cancer and other conditions. The PROMIS initiative is covered in more detail by other presenters.

More recently, the NCI Symptoms and Quality of Life Steering Committee have been working to identify core outcomes to measure in cancer clinical trials, beginning with a focus on symptoms (with HRQOL and functional status to be addressed later). Through a review of the literature and analysis of a variety of data sources, they identified a candidate core set of symptoms for use across cancer types and for three specific cancers (prostate, ovarian, and head and neck). This effort focused primarily on the outcomes to measure, rather than the measurement instruments.

Key lessons from the cancer community's search for core measures are: (1) it is difficult to identify a universally applicable measurement instrument and each study's particular objectives will drive the measurement model; (2) identifying the core domains or concepts to measure is an important first step; (3) identifying core measures for particular applications (e.g., clinical trials) may be more feasible; and (4) methodological advances are improving our ability to compare outcomes measured with different instruments.

**The PROMIS Experience**

*Kevin Weinfurt, Ph.D., Professor, Department of Psychiatry and Behavioral Science, Duke Clinical Research Institute, Duke University*

Dr. Weinfurt stated that the PROMIS PRO measurement system sought to include a set of measures for adults and children with the types of domains affected across chronic diseases, advance the science of PRO, and develop a software set.

An item bank is a collection of items to measure a single domain. Any item can be used to provide necessary information. An item bank was developed through a dynamic process of qualitative research, testing, analysis/interpretation, and refinement of results.

The PROMIS framework is advantageous because it can be used in varying situations, trials, and diseases. Recommendations for measuring patient-reported experiences of LUTS are: (1) be sensitive to different domains of clinical research versus clinical care; (2) where appropriate, use IRT; (3) carefully plan so that the entire severity range is sampled; (4) conduct many cognitive interviews; and (5) study the recall period (i.e., what patients will and will not remember in the past week or month).

**Emerging Technologies in the Measurement of PRO**

*James W.Griffith, Ph.D., Research Assistant Professor, Department of Medical Social Sciences, Northwestern University; Senior Visiting Fellow, KU Leuven, Belgium.*

Dr. Griffith, a clinical psychologist, explained several aspects of questionnaire development, including creating content that asks the right questions (a.k.a. content validity), being inclusive so to maximize participation from those who might otherwise have difficulties in completing questionnaires, creating tools that are accepted by patients and clinicians, and being both brief and precise. Self-report questionnaires serve many purposes including measuring treatment outcomes, phenotyping and diagnosing patients, and screening for medical conditions. The purpose of a questionnaire often helps drive the creation of the content and delivery system.

LUTS can be organized into many potential dimensions including urgency, frequency, nocturia, hematuria, and can also co-occur with other symptoms such as erectile dysfunction, pain, anxiety and depression. Assessment of multiple dimensions is an emerging development that can assist in the description of complex diseases, such as major depression. For example, in phenotyping patients, two depression sufferers can have zero overlapping symptoms and yet receive the same diagnostic label. Solving this conundrum can be achieved by organizing the phenotype into different symptom dimensions, such as in David Watson's quadripartite model that includes dysphoria, lassitude, suicidality, insomnia, and appetite loss/gain. Although some of these are relatively depression-specific (e.g., lassitude), others are found across other conditions. Detailed phenotyping can assist in the separation of patients suffering from depression versus other disorders, and may help to plan treatments. Assessment of treatment outcome and patient phenotypes can both be multidimensional, and using well-validated and potentially computer-adaptive questionnaires can assess each dimension in a short amount of time, providing rich information and a low cost.

Computerization is increasingly influencing PRO measurement with the advent of computer-adaptive testing and the use of mobile devices and tablets to collect health data. Dr. Griffith stressed that technology does not alter the fundamental work performed to create high-quality content. Thus, a state-of-the-art questionnaire includes front-end scholarly work including literature review, interviews with patients and clinicians, and quantitative testing to develop content, and then subsequent work to deliver this content in a way that is efficient and user-friendly.

Given the existing diversity of LUTS measures, Dr. Griffith concluded by discussing a PROsetta stone project" in which different instruments can measure the same construct can be linked on a common metric. This research might be applied to LUTS measures, where there are many different instruments (and potentially news ones being developed), but a need to have common and easy-to-interpret units of measure.

**How Best to Validate a Symptom Measurement Tool for LUTD?**
*Stephen Van Den Eeden, Ph.D., Senior Epidemiologist, Division of Research, Kaiser Permanente Northern California*

In validating a symptom measurement tool for LUTD, Dr. Van Den Eeden stated that the use must be understood (e.g., clinical trials, clinical research, epidemiology). Additionally, the phenotype must be defined (e.g., etiologies/mechanism), so as to sort out the contributions of prostate obstruction, bladder dysfunction, infection, pelvic floor disorders, and so forth on LUTD.

Differentiating between phenotypes can be achieved via multiple questionnaires, domains with a single questionnaire, PROMIS-like item banks, and/or modules. There is still a struggle to attain a phenotype "gold standard."

When a new survey is developed, it is important to test. This will assist in the identification of valid populations for its use and representation of appropriate age, gender, race/ethnicity, education, and other SES characteristics. The mode of administration also should be considered (e.g., self, staff, computerized).

**New and Innovative Phenotyping Methods in Male LUTS**
*Kevin McVary, M.D., Professor, Department of Urology, Feinberg School of Medicine, Northwestern University*

The question of what influences the rate of prostate growth in different men is not yet understood. This uncertainty partially is due to variability in patient responses on IPSS surveys and complex phenotypes for LUTS patients.

Clinical attributes for phenotyping include age, which groups patients effectively, prostate volume, and PSA levels. It is not clear, however, if grouping patients in these ways is useful. Prostate size and other proxies are potential phenotyping attributes, but symptoms are optimal for phenotyping patients. Biometrics, such as inflammation and serum C-reactive protein (CRP) levels are worthwhile avenues for patient phenotyping.

Regarding comorbidities, there is an association between metabolic syndrome and LUTS in men, especially younger men. LUTS increases with elevated autonomic activity. More severe LUTS is reported with increasing sleep disturbance in men, and this is not driven by nocturia.

Genetic sequence variants (i.e., single nucleotide polymorphisms [SNPs]) are associated with LUTS. In genome-wide association studies, there are a number of SNPs that associate with prostate cancer risk. There also is an increased risk of undergoing TURP associated with certain SNPs. Additionally, there is one SNP that has shown a nominally significant association with the BPH phenotype.

Dr. McVary concluded by saying that there are clinical uses for SNPs in this new era of genotyping.

**What Are the Barriers to Moving Forward?**
*John Wei, M.D., Professor Urology, Department of Urology, University of Michigan*

In the measurement of LUTD it is necessary to progress research quality, clinical usefulness, standardization, FDA/regulatory compliance, and LUTD phenotyping.

The first significant barrier is disagreement among colleagues regarding whether the necessary tools already have been developed or not, if tools should be made for research or clinicians, and so forth. A second barrier is disagreement on the domains that should be included in LUTD (e.g., BOO, LUTS, OAB, detrusor failure, pain, UI, sexual issues, bother, adaptation, expectations). Deciding the optimal approach for undertaking PRO is a third barrier. There is no consensus whether more validated condition-specific instruments should be created, existing instruments should be merged, or an item bank should be created to allow meta analyses. Two final barriers include if PROs should be extended (e.g., clinical or observer ROs) and where the "line" should be placed regarding these assessment tools.

***Discussion***

Another barrier to consider is how to encourage fellow researchers and clinicians to use the tools that are devised. Although there are thousands of incontinence surgeries annually, the largest studies that report outcomes use merely hundreds of patients. This is a small fraction of those who undergo operations, and so it should be made clear if assessment tools are made for research and/or clinicians. There is no consensus on how to best measure success after incontinence surgery. A symptom composite is not a diagnosis. There are significant barriers, but the LUTD community must overcome them.

Creating item banks and undertaking item response theory can be time consuming initially and is a complex process when it involves utilizing, developing, and testing extant items. Certain concepts are amenable to latent variable modeling, while others are less so. IRT requires large samples because many statistical parameters are estimated in IRT-based analyses. These are challenges to the development of item banks using IRT.

The reason for the MOMUS meeting was to have an opportunity to create system innovations and restructure, develop a conceptual model for LUTD, discuss the inclusion of phenotyping, and detail the endpoints that should be measured from patient, clinical, and observer perspectives. The current conceptual model is not a statistical model but is a description of what LUTS community members think. It is imperative for researchers and clinicians to "break down walls" and redesign the LUTD system. Inherent in this is a new emphasis on talking and listening to patients, which can be achieved via focus group discussions. It is important to identify core concepts that help differentiate urologic symptoms from other related disorders.

Overactive bladder should score as frequency and urgency on a questionnaire, but behaviorally people will frequently urinate to avoid feeling urgency. This creates a scoring problem. New PROMIS methods should be able to overcome this situation and take such issues into account for

a reliable composite score. Additionally, an all-inclusive versus parsing survey problem exists. Symptoms should be understood, but the underlying etiology is imperative to know.

## BREAKOUT SESSIONS

### Session I: What is Missing in Current Measurement Tools for Male LUTD?

*Moderator: Marcus Drake, M.D., Ph.D.*

LUTD conditions are best defined by symptoms rather than a supposition of a mechanism. This concept makes the need for symptom clusters necessary. Current measurement tools and scientific discussions have organs identified in LUTS names, which often suggest an inappropriate treatment pathway.

LUTS that should be included in current measurement tools for male LUTD are storage, voiding, post-micturition, pain, and incontinence. Storage symptoms should include leakage, frequency, urgency, storage pain, and nocturia. Leakage is not measured by IPSS and can be stress, urge, continuous, or without-awareness leakage. Urgency and storage pain are not measured by IPSS and are likewise important. Nocturia should have a high-volume and low-volume differentiation.

Voiding symptoms that should be measured are poor flow, hesitancy, voiding pain, and terminal dribbling. Both terminal dribbling and voiding pain are not currently measured by IPSS, but should be included. Voiding pain has supra-pubic, urethral, and perineal subcategories for further symptom identification.

Post-micturition means that there is a sense of incomplete emptying, pain, and/or dribble (not indicated by the IPSS). If these are present, post voiding residue (PVR) should be considered. The differentiation between dribble and leakage was discussed (i.e., urine loss immediately following emptying versus urine loss during regular daily activities) and it was determined that dribble is appropriate for post-micturition situations.

Measurements that cannot be overlooked are “normal state” indices. “Normal state” implies that the patient has taken no medications and not undergone surgery, has non-bothersome symptoms, and has no activity restrictions for urologic reasons. This measurement state is imperative because for epidemiology and pre-and post-treatment, thoughtful and understanding dialogue is necessary to ascertain the key indicators that define a baseline. This allows for meaningful information to be taken from the study, and valid, well-formed conclusions to be developed.

Other measurements that are sometimes missing, but should be included are symptom modifiers such as behavior adaptations. This refers to fluid adjustment, activity level (sweating out extra fluid through exercise versus avoidance of exercise), toilet mapping, and containment methods. Symptom cofactors are important and should comprise the following co-occurring symptoms: sexual function, bowel function, generalized pain, sleep disorders, and fatigue.

As discussed, key states and conditions that measurements should encapsulate are “normal” and “abnormal” (e.g., obstructing prostate, benign prostatic obstruction [BPO], OAB, LUTS without

obstruction [e.g., DUA], prostatitis, and PPUI). Avoiding clinical pathway bias can be achieved by focusing on the aforementioned symptom clusters.

It was concluded that the result of including the appropriate measurement tools will have the impact of patients achieving the ability to live with their symptoms, which is preferable to bother. Tools will need to capture storage, voiding, post-micturition, and adaptation symptoms.

**Session II: What is Missing in Current Measurement Tools for Female LUTD?**
*Moderators: Michael Albo, M.D. and Lisa Begg, Dr.P.H.*

Measurement tools for female LUTD are used for a variety of purposes, including clinical practice and research. The discussion focused first on clinical tools, primarily patient reporting and objective assessment that measure LUTD symptoms including those instruments that also connect to biology and other measurement tools. Many tools exist today that measure LUTD in females, but participants were asked to consider what is missing in the available LUTD measurement tools.

The majority of female symptoms originate with storage issues such as incontinence, frequency, urgency, and nocturia. Currently, an effective technological method of measuring sensation is missing, and patient questionnaires do not capture bladder sensation information successfully. Storage, urgency, and pain symptoms often can be assessed with a patient questionnaire, but bladder changes that precede, lead to, or differentiate such symptoms remain difficult to measure; patient symptoms must be better synchronized with actual measurements of biological events in the bladder (e.g., when does the bladder physiologically change and LUTD symptoms begin to occur?).

Participants agreed that many LUTD symptom questionnaires exist today, and most address incontinence questions, such as how often a patient urinates, the amount of leakage a patient is experiencing, and when leakage most often occurs (e.g., during pressure, stress, or urgency). Currently, there is no global questionnaire available for clinical use, and existing questionnaires do not capture female patients' LUTD symptoms (pain, incontinence, urgency, and frequency) comprehensively. In addition, questionnaires such as the AUA assessment tool do not effectively measure the symptoms that patients most commonly present during clinical visits.

In addition, inconvenience and bother remain the major concern of most patients, but few measurement tools ask patients to identify their most bothersome symptom or the primary reason for medical visits; also, few questionnaires capture patient goals or expectations, or measure QOL. Revisions to current intake forms could include a simple symptom rating system that patients could use to identify their most troublesome symptoms. Baseline information also should incorporate goals, expectations, current QOL, and the extent of inconvenience that patients are experiencing at the time. Such information would allow clinicians to assess patient motivation better and lead to improved outcomes.

Clinicians must consider how questionnaires will be used clinically (e.g., how such assessment tools will impact clinicians' decisions). If a clinician's primary goal is to improve patient

symptoms, then a global questionnaire is needed that can demonstrate change by helping patients' symptoms improve over time and facilitate achievement of patients' goals. Research is needed that translates into methods to eliminate symptoms in large numbers of patients; statistically significant results likely are meaningless to patients who are experiencing LUTD symptoms.

Participants agreed that a revised clinical symptom questionnaire should: (1) provide increased opportunities for clinicians to discern patients' chief complaints; (2) improve clinicians' abilities to identify the reasons that patients visit a urologist; (3) increase dialog between patients and clinicians concerning patient goals and expectations; and (4) be responsive to change. Treatment must demonstrate outcomes, and the goal of revised measurement tools should facilitate those outcomes.

Feedback from patients on which questionnaire(s) most accurately captured their chief complaints, symptoms, and QOL may benefit clinicians and help with the development of a global measurement tool for females with LUTD. Patient feedback also would provide clinicians with an opportunity to assess the language skills of patients using the questionnaire because inadequate reading or vocabulary skills will impact responses.

For research purposes, questionnaires initially are used to screen patients based on inclusion/exclusion criteria according to the hypothesis being tested. Research questionnaires, therefore, are targeted based on the necessary study population and likely will differ according to research needs. Questionnaires also will differ at baseline, 6 months, 12 months, and so forth, and will be tailored for specific populations and hypotheses in a research study.

Research measurement tools to assess female LUTD symptoms should provide investigators with a better understanding of frequency categories and the history and physical backgrounds that created frequency symptoms. Currently, patient symptoms often are "lumped together," and questionnaires do not allow patients to be differentiated, which ultimately hinders research, patient treatment, and outcomes.

In addition, research questionnaires that are inclusive and can measure and rank symptoms based on how bothersome they are to patients would have a positive effect on treatment outcomes in research. Measurement tools also should be harmonized and strategies be developed to interrelate the various questionnaires to ensure similar data (e.g., symptoms and populations are compared). Tools also must include goals, achievements, and outcomes.

## Session III: How to Validate the New Measurement Instrument

*Moderators: Kevin Weinfurt, Ph.D. and Stephen Van Den Eeden, Ph.D.*

### *Context*

The validity of an instrument refers to the degree to which the instrument measures what it is intended to measure. There are different types of validity and different methods one can use to ascertain validity. To determine the types and methods of validation appropriate for a given

measure, it is essential to specify the context within which the measure will be used. The context of use includes (1) the intended population and (2) the purpose for the assessment. We address each in turn.

The *intended population* for an instrument for LUTD could be anyone suffering symptoms of the lower urinary tract. More specific subpopulations are possible, however, such as males versus females. For purposes of discussion, our group assumed that the instrument would be targeted toward all LUTD sufferers, regardless of sex, but that efforts would be required to evaluate the possibility that the instrument is differentially valid for males and females.

The *purpose of the instrument* could be classified in terms of either research or clinical practice. Within the research category, there are further distinctions that would be relevant to evaluating validity. An instrument could be intended for use as an endpoint in randomized clinical trials and, perhaps, as the basis for a labeling claim in a regulatory context. Alternatively, the instrument could be intended for use as an outcome used in epidemiological studies. In the arena of clinical care, there are also multiple purposes for an instrument. The instrument could be used to facilitate and enrich the diagnostic process by providing a more comprehensive, systematic collection of symptom data. Alternatively, the instrument could be used in clinical care to measure responses to treatment over time.

A complete discussion of validity would require careful elaboration of the standards and methods for validity assessment in each of the contexts described above. For the remainder of our discussion, we focused mostly on research applications.

### *The PROMIS™ Experience*

The National Institutes of Health's Patient-Reported Outcomes Measurement Information System (PROMIS) Network (www.nihpromis.org) has developed frameworks that could be helpful in charting a course for the development and validation of a patient-reported measure of LUTD. The PROMIS Network is making generic tools for measuring patient-reported health for aspects of health ("domains"), such as pain, physical functioning, and fatigue, that are relevant to many chronic diseases. PROMIS uses a mixture of qualitative and quantitative methods to examine validity throughout the development process. The qualitative methods include focus groups and patient interviews to define the domain of interest and to ensure that the final measure addresses all aspects of the domain under study—a characteristic known as *content validity*. Cognitive interviews are used to further ensure that patients understand items in the way the developers' intended them to be understood, which is another requirement for a valid instrument.

Quantitative techniques rely on so-called classical test theory and item response theory (IRT) methods to examine whether the items behave as they should, if the instrument measures what it is intended to measure. An important analysis made possible by the use of IRT is *differential item functioning*, which can be used to determine whether an individual item behaves the same way for different groups. Our breakout group identified age and sex as possible variables that

would define different populations that should be evaluated in terms of differential item functioning.

The PROMIS Network recently completed a set of studies that examined the responsiveness of the PROMIS measures in different clinical populations. Such longitudinal designs using populations of patients who undergo change are critical for evaluating the validity of an instrument for measuring LUTD. The measure is valid to the extent that patients who undergo significant clinical improvements or worsening show corresponding changes in their scores on the new instrument. Such longitudinal clinical studies can also be used to examine the relationship between scores on the patient-reported instrument and values of "objective" clinical measures (e.g., lab values). If there is no relationship at all, this calls the validity of the patient-reported measure into question. If the relationship is too strong, however, it calls into question the utility of the patient-reported measure. Typically, the patient-reported measure is required because there is a concern that the objective clinical measure does not tell the whole story. A perfect correlation between the patient-reported and objective measures suggests that nothing is being added to the story by the patient-reported measure.

### *A First Step: Evaluating the Content Validity of Existing Measures*

The group discussed the fact that many measures of LUTD symptoms exist, although some are intended for use in specific subgroups, e.g., women with urinary incontinence, men posttreatment for prostate cancer, men with LUTD from benign etiologies, etc. There is also the well known AUASI/IPSS measure. It would be useful to know to what degree existing items address the totality of experiences of LUTD sufferers. One avenue for examining this would be to pool items intended to measure a specific domain, such as urgency, from these various measures and sort them into subdomains. It would then be possible to elicit data from a variety of LUTD sufferers via focus groups or individual interviews to create a comprehensive list of their symptoms relevant to the domain (e.g., urgency) and the effects of those symptoms on their daily lives. Finally, one could determine how well existing items addressed all of the issues raised by patients during the qualitative interviews. This analysis would highlight aspects of the domain that are not well captured by items from existing measures, pointing the way for further item development.

The question of how best to validate a new measure is a complex one that we were not able to address in sufficient detail in the context of this meeting. Our discussion did, however, highlight several important considerations and opportunities that could advance the science of assessment for patients with LUTD.

## REPORTS OF THE BREAKOUT SESSION LEADERS

### *Session I: Marcus Drake, M.D., Ph.D.*

Key symptoms and symptom clusters are important in measurement tools for male LUTD. Conditions are best defined by symptoms rather than by the organ. For example, OAB has the organ "bladder" in the name, which is inappropriate because many other conditions cause urgency as well. Deriving and grouping symptom clusters that are important is a key point. These clusters should be post-micturition, storage, voiding, pain, and incontinence.

The post-micturition cluster includes a sensation of incomplete emptying, dribbling, and incomplete emptying without sensation. The storage cluster involves leakage, frequency, urgency, pain (storage), nocturia (high or low volume is essential for informing in the next clinical step to be undertaken). The voiding cluster encompasses poor flow, hesitancy, pain (with location establishment), and dribbling at the end of voiding. The final cluster is the normal state, which includes no medications or surgeries. The important aspect is to have a low level of preceding symptoms, no bother, and the lack of symptoms cannot be due to medication consumption. The individual should also not feel limited in activities undertaken (i.e., no avoidance).

Symptom modifiers are behavioral adaptations, such as fluid adjustment, activity level, toilet mapping, and containment methods (pad counts). Symptom cofactors include sexual function, bowel, generalized pain, sleep disorders, and fatigue.

The key states and conditions are normal and abnormal. Abnormal encompasses BPO, OAB, LUTS without obstruction, prostatitis, PPUI, and an obstructing prostate. The measurement tool should avoiding biasing clinical pathways by focusing on symptom clusters.

### *Session II: Michael Albo, M.D. and Lisa Begg, Dr.P.H.*

It is important to determine the clinical versus research use of a measurement tool. Clinicians are motivated because a patient is complaining of symptoms. The symptoms and their impact is what drive patients into the office. Clinicians need to understand patient goals, expectations, and bother, and to not clump symptoms together for treatment, but to treat the symptom that is bothering the patient. Clinical measurements are needed to monitor treatment effectiveness and to identify a diagnosis.

Regarding research, many validated questionnaires exist for females that measure aspects of pelvic floor dysfunction and so forth. A concern is that many such tools are available and they need to be harmonized for data comparisons.

A questionnaire should identify the most bothersome symptom and the symptom that is motivating a patient to visit a doctor.

***Session III: Kevin Weinfurt, Ph.D. and Stephen Van Den Eeden, Ph.D.***

Determining how best to validate a new measurement instrument is challenging because the new instrument is as yet unknown. The validation process and data collection depend upon the instrument purpose (e.g., measuring trial endpoints, diagnoses, response to treatment).

There are many measures of certain domains, such as urgency, and instead of developing a new measure, there should be improved understanding of the currently available tools and their validity. A validation process similar to PROMIS, in which there are domains and sub-domains that bin items according to commonality, is reasonable. There is an emergent conceptual model for domains.

Confidence in measurements matching what a patient feels as urgency, for example, can be determined via patient-focused group discussions. These discussions can labor under the concept of determining the subcomponents that are endorsed by patients due to their relevance and importance. This is a validation process that will ensure the items appropriately cover the concept being measured. New items could be developed and old items revised based on cognitive pool interview results. Expert review would also assist in honing appropriate items.

There is a need to more systematically relate what currently is available with what patients are reporting.

***Session IV: John Wei, M.D. and Kevin McVary, M.D.***

Determining the optimal method to phenotype patients with symptomatic LUTD is complicated because the exact definition of phenotype is complex. It is certain that measuring symptoms alone is not adequate, and biomarkers and other physiology methods are required. To resolve these intricacies, patient profiles should be parsed into information pieces. These pieces should include symptoms, physiological parameters, genotype, and biomarkers (e.g., prostate size, pain).

Parsing patient profiles into information packets in this fashion will allow the data to “speak for themselves” by undergoing bioinformatic analyses (e.g., cluster analyses, MDA). This methodology will confirm what is already known and important, and should also identify new phenotypes that have been heretofore unknown or atypical.

**WRAP-UP**

*Robert Star, M.D., Division Director, Division of Kidney, Urologic, and Hematologic Diseases, NIDDK, NIH*

Dr. Star thanked conference participants for an outstanding meeting. He also acknowledged the organizing committee, speakers, and meeting contractor for their hard work. He commented that there is inherent difficulty in challenging the paradigms that drive the field of Lower Urinary Tract Disorders. While discussions were spirited and disagreements occurred, the breakout groups rose to the challenge of considering how to move the new field of Lower Urinary Tract Dysfunction (LUTD) forward.

He summarized the discussions from several perspectives:

From a research perspective, there are many available instruments, but the content validity of these instruments is uncertain. Speakers objectively discussed the shortcomings of the available instruments; i.e., that current symptom scores are misleading in clinical settings and correlate poorly with patient satisfaction. Clinical phenotyping is primitive and does not identify useful and positive information [e.g., what organ is involved?, what is the underlying disease process(s), what targets should be considered].

From an FDA perspective, how patients feel, survive and function is critical. They are developing new guidances on Patient Reported Outcomes, and the instruments used to assess them. The context validity and context of use is extremely important. Developing and qualifying instruments for diagnosis, stratification, and outcomes is key. No one tool is expected to accomplish everything.

From a clinical perspective, listening to patients' complaints is imperative. Instruments often do not capture important areas, or they capture them incompletely. Current instruments may assist in refining a clinician's impression of a patient, but there is dissociation between a change in measurement tool score and what a patient articulates. For example, bother and exacerbation are currently not well measured. This disconnect impedes progress of the LUTD field in the clinical setting, and also in the research setting.

From the clinical trial and epidemiological perspective, patients are heterogeneous and difficult to compartmentalize. Connecting symptoms to an underlying pathophysiology is exceedingly difficult; new conceptual models and new targets for intervention are quite rare in the LUTD field.

To move the LUTD field forward, barriers to thinking and old paradigms must be dismantled. There is a great need to listen anew to patients about their symptom complexes, what bothers them, probe for symptom flares and remissions, and understand how the patients adapt over time. New conceptual models are needed that integrate temporal changes in symptom complexes (including flare, bother, adaptation, remission and so forth) with physiology and pathophysiology (including structure, biomechanics, function, genetics, environmental influences), and remission/repair/adaptation pathways. Better phenotyping tools are needed, and confounding factors and symptoms should be identified. All of this must be integrated to discover better

interventions, be they pharmacologic, structural, electrical, or behavioral. For the field to make substantial progress, all of these critical components must be better understood.

He concluded by saying that NIDDK is greatly interested the new field of Lower Urinary Tract Dysfunction, and will soon be announcing an important initiative to catalyze the LUTD community to begin to accomplish what has been discussed at this meeting. It is not only a critical time, but the correct time for the LUTD field to begin to make progress on these scientific opportunities.

Dr. Star adjourned the meeting at 12:45 p.m. EST.