Validation of Bladder Health **I**nstrument for **E**valuation in **W**omen **(VIEW)**

A Protocol for the Prevention of Lower Urinary Tract Symptoms (PLUS) Consortium

Protocol Committee:
Melissa Constantine, University of Minnesota
Emily Lukacz, University of California San Diego
Todd Rockwood, University of Minnesota
Alayne Markland, University of Alabama at Birmingham
Elizabeth Mueller, Loyola
Jerry Lowder, Washington University
Diane Newman, University of Pennsylvania
Leslie Rickey, Yale
Kyle Rudser, University of Minnesota
Lisa Kane-Low, University of Michigan

Study Title	**V**alidation of Bladder Health **I**nstrument for **E**valuation in **W**omen **(VIEW)**
Study Design	Prospective observational
Primary Objective	To assess the reliability and validity of a Bladder Health Instrument (BHI) for measurement of bladder health among women across the life-course
Secondary Objective(s)	To evaluate effects of mode of administration of the BHI across paper and pencil versus electronic
Research Intervention(s)/Investigational Agents	N/A
IND/IDE # (if applicable)	N/A
Investigational Drug Services # (if applicable)	N/A
Study Population	Adult English fluent women
Sample Size (number of completes)	Total: n=1202 General population sample: n=694 Clinical evaluation sample: Asymptomatic & symptomatic: n=354 Post-partum sample: n=154
Study Duration for Individual Participants	**General population:** < 2 hour for BHI and survey completion (electronic or written) with up to six weeks for a subset of participants who complete both the initial validation questionnaire and those who also complete the re-test questionnaire for the reliability evaluation and 2 days documentation of symptoms on voiding diary. **Clinical sample:** Up to 8 weeks for completion of BHI (< 2 hr total) in addition to 3 days documentation of voiding (2 days symptoms + 1 day volume frequency) diaries prior to in-person visit, plus < 3 hour in-person visit including paper towel testing, uroflowmetry, post void residual and judge evaluation.

REVISION HISTORY

Version #	Date	Summary of changes from prior version	Consent Change?
Version 0.2	6/21/2018	Clarification of clinical evaluation tests Further revisions per team input Revised Table 1	
Version 0.3.1	7/5/2018	Addition of Figures 1, 4 and 5, revision of Figure 6. Creation of a Measures section. Further explanatory text in Analysis section.	
Version 0.4.1	9/24/2018	Addition Logistics survey administration, Re-formatting.	
Version 5.2	12/26/18	Incorporate all changes, edits and comments from 5.1	
Version 1-8-19-v1	1/11/19	Expand bladder judge inclusion criteria Remove retest for clinical eval sample Inclusion of "hat" with volume-frequency diary	
Version 6.0	3/20/2019	Edits per EEP comments including: additional language for re-test to evaluate bladder function change since completion of BHI. Clarifying text re: CASI mode in general population sample Inclusion of at-home self-administered PTT Addition of bladder cancer dx or history as exclusion criterion Further details provided for gen pop sampling and clinical eval sampling	
Version 6.1	4/29/2019	Exclusion of at-home PTT Additional exclusion criteria added for clinical eval sample Split Figure 4 into 4a and 4b	
Version 6.2	June 4, 2019	Clarification of inclusion/exclusion criteria	
Version 6.3	June 14, 2019	Addition of randomization for NIDDK Data Repository language	
Version 6.4	June 17, 2019	Clarification of randomization for NIDDK Data Repository language	
Version 6.5	July 25, 2019	Clarification and more thorough details of Clinical testing	

I. Purpose

The Prevention of Lower Urinary Symptoms (PLUS) Research Consortium is working to optimize prevention of lower urinary tract symptoms (LUTS) in women and adolescent females across their life spans. The ability to measure bladder health and key risk and protective factors is crucial to the PLUS mission. To describe and measure the spectrum of bladder health in diverse populations, researchers need a valid and reliable instrument. To date, the Consortium's work on design of a bladder health instrument has been a culmination of expert opinion, information from focus groups, and incorporation of previously validated items and language where appropriate, along with cognitive interviews of participants from the general public. The next step in the consortium's work is to prospectively collect data to test and validate bladder health instrument (BHI) items for inclusion in a final bladder health scale (BHS) that can assess the full range of bladder health of women.

Consistent with the World Health Organization's definition of health[1], the PLUS Consortium conceptualizes bladder health as *"a complete state of physical, mental and social well-being related to bladder function, and not merely the absence of LUTS[2]"* with function that *"permits daily activities, adapts to short term physical or environmental stressors, and allows optimal well-being (e.g., travel, exercise, social, occupational or other activities)."* The consortium came to consensus regarding the fundamental conceptual framework that would guide development.[3] Healthy bladder function encompasses storage, emptying and bio regulatory functions of the bladder. One of the goals of PLUS is to create a measure that can assesses the spectrum from very healthy to very unhealthy within a community dwelling population across the life course, necessitating a reliance on self-administered measures. As shown in Figure 1 (panel A), bladder health that can be assessed using survey methodology is conceived of three concepts: storage, emptying and psycho-social. LUTS plays a central role relative to identifying individuals along the spectrum, and therefore is integrated into the measurement model (Figure 1, Panel B). The measurement of bio regulatory function will not be feasible without specimen collection and as such, this element of bladder health is not measured in the BHI. To note, until validation and reliability evaluation are complete, the bladder health items are referred to as a bladder health instrument (BHI). Following full evaluation, the expectation is a resultant bladder health scale (BHS).

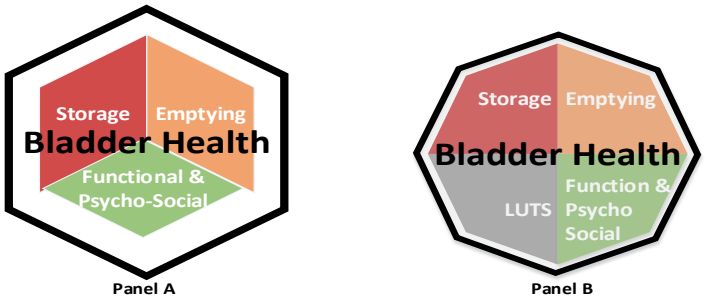


Figure 1. Conceptual Framework for Bladder Health measurement

The BHI is being designed to have a life course perspective and the ability to capture transient symptoms at various life stages that may affect current and future bladder health. To this end, the current protocol embarks on initial validation processes in adult women, with secondary emphasis of including postpartum women, recognizing the high prevalence of LUTS and potential for intervention in this population. In subsequent studies, the BHI will be refined to assess bladder health in adolescent and Spanish speaking women.

II. AIMS

Primary Aims

- **Aim 1. To develop and establish internal and external validity of a bladder health instrument (BHI) for use in population-based research.**
- **Aim 2. To develop and establish internal and external validity of a bladder health instrument (BHI) for use in clinical research.**
- **Aim 3. To develop and establish internal and external validity of a bladder health instrument (BHI) for use in a post-partum population.**

Measuring bladder health is central to all future work of the PLUS Research Consortium. As such, a main goal is to create and validate a BHI that is focused on assessing degrees of bladder health among an ethnically and geographically diverse population of community-dwelling women. In part, content validity for the BHI is currently being established based on expert opinion, adaptation of condition specific instruments, incorporation of language from prior PLUS studies involving focus groups and cognitive evaluation. The proposed research will extend this internal validation of the BHI by evaluating dimensional validity of the BHI. A further goal is to establish an external validation with a level of evidence that will allow valid inference to be made based on BHI score in both population based and clinical based research of community-dwelling women and post-partum women. The finalized instrument will be the end product of this study with internal validity (content, construct and dimensional), and external validity (criterion) established for the BHI.

Secondary Aim:

Aim 4. To evaluate the effect of mode of administration (paper and pencil vs. web-based) on item distributions, response rates, and validity.

The BHI is designed to be a self-administered questionnaire (SAQ). At the current time, paper and pencil (PAPI) and web-based/computer assisted self-administered (CASI) instruments are the primary modes the consortium is focusing on for validation of the BHI. There is an accumulation of evidence and knowledge of context based effects associated with mode of survey administration and the design of surveys for mixed-modes.[4–7] With rapidly changing technology and increasing availability of online survey administration, the ability to optimally capture data is paramount. While some may prefer the PAPI administration, it is the

expectation that younger populations of women and adolescent females will prefer and be more apt to utilize the CASI administration. As such, assessment of the BHI in both modes is critical to assess the validity by mode of administration as well as to evaluate survey completion rates based on demographic characteristics. These data will optimize recruitment and survey administration methods for future population based cohort studies.

Some scales have shown that while there are mode effects, the underlying psychometric properties of multi-dimensional scales can remain constant between modes.[8,9] Other potential biases or sources of error associated with mode of administration are unknown. While a computer mode may increase efficiency of questionnaire completion, it also may produce measurement error; use of a computer may alter the distribution of BHI scores.[8,9] However, where any change in distribution is found to be constant, this effect can be controlled by score calibration. Therefore, a mode experiment to identify and/or quantify any effect of mode of administration is included in the study protocol.

III. PRELIMINARY DATA

The process of developing and validating a measurement instrument is iterative. Generally, the sequence of tasks for the BHI development is:

1. Development of conceptual model with well specified constructs (or dimensions)
2. Generation of an item pool of new or existing items mapped to constructs
3. Cognitive evaluation of items (face/content validity)
4. Reliability and validity analyses
5. Responsiveness, evaluation of sensitivity to change
6. Establishment of validity in special populations (Spanish fluent and adolescents)

Figure 2 (below) depicts the consortium's overall plan for development and validation of a bladder health measurement scale (BHS). Steps 1 and 2 have been completed. Step 3 is expected to be completed by May 2019. Step 4, the proposed validation work of this protocol, will commence by June 2019 and be complete by spring 2020. Steps 5 and 6 are planned to be completed through future protocols.

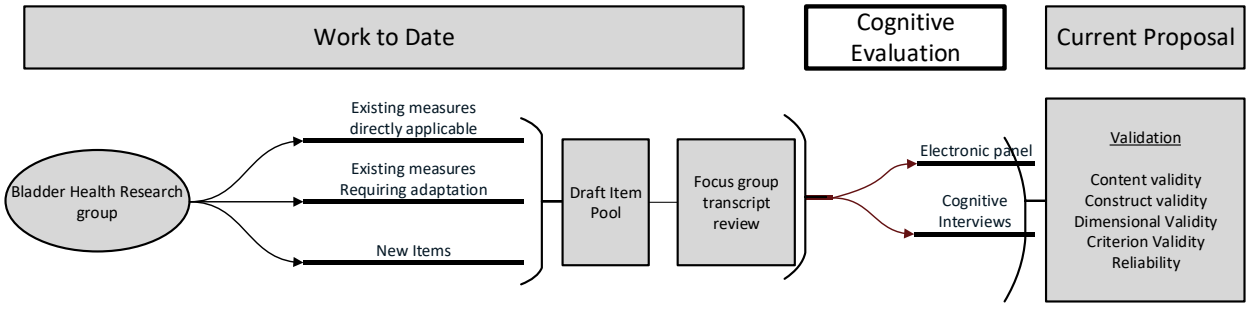


Figure 2. Schematic of the overall plan for a bladder health instrument development and validation

III.A. Conceptual framework and item pool

Per the approved concept proposal for the development of an overall measurement tool, tasks 1 and 2 have been completed. Of note, the model of bladder health has been reviewed using community engagement approaches with community partners at PLUS research centers to ensure that it is understandable and comprehensive of lower urinary tract function and its impact on daily activities. The BHS is intended to assess bladder health across the life course, identifying current bladder health status at a single point in time. An additional objective of the BHS is to measures changes in bladder health over time. While this protocol is not designed to assess sensitivity to change, future work of the consortium intends to examine the validity of the instrument for longitudinal measurement.

Preliminary work in development of the BHI has included review of existing symptom scales and drafting novel items to assess the elements of our conceptual model of bladder health, which included the following domains: Storage and Emptying symptoms, as well as functional and psycho-social impacts of bladder function. A literature search was conducted which identified LUTS symptom scales and measures as well as condition specific quality of life. Virtually all of the existing instruments were designed to measure LUTS in an affected clinical population and assess impact of treatment. Therefore; language, time frame, and response categories were modified in order to capture the spectrum and range (very healthy to very unhealthy) of bladder health in the population that will include women with absent, mild, moderate, and severe LUTS. Review of these items was conducted by consortium members with subsequent refinement of items that focus on health aspects of bladder function. The item pool has been further refined based on complete review of focus group transcripts of women enrolled in a prior study conducted by PLUS called SHARE. Focus groups were conducted with group participants grouped into the following age groups: 18-24, 25-44, 45-64, and 65+. The transcripts were reviewed by 2 independent investigators, both with measurement expertise. Following individual review, the two investigators shared general impressions to identify commonality and divergence in pereceptions of gestalt of the focus group conversations as well as comparison of items each investigator marked for re-writing or deletion and suggested new items. A lengthy adjudication process was employed, with written justification provided for item revision, deletion or addition. From this process, a preliminary set of approximately 85 items was developed for evaluation with an on-line electronic panel sample of women (e-panel) and cognitive interview testing. The BHI is also currently undergoing refinement with cognitive evaluation as described below.

III.B. Cognitive Evaluation of Items

The cognitive evaluation of items was conducted in the PLUS Consortium's Clarification of Language, Evaluation And Refinement of questions (CLEAR) study. There are two general sets or types of items within the BHI. The first set of items are those that every respondent will be

presented with. These items are intended to capture respondent perception of their overall bladder health and impact of their bladder on their life. These "universal" items are answerable for every woman independent of the presence of LUTS. The second set of items are designed to capture information regarding presence of LUTS, including impact on activities and quality of life. Therefore, these symptom specific items are only asked of respondents who indicate experiencing one of the 6 LUTS domains (urinary urgency, frequency, incontinence, voiding dysfunction, urinary pain and urinary tract infections). There is a plethora of validated instruments designed to measure severity of LUTS and impact on quality of life, providing a rich pool of item structures to inform the BHI symptom specific item development. Conversely, there is a dearth of validated instruments that measure the constructs or dimensions of bladder health that the BHI universal items intend to measure. While less is known about how women will interpret or respond to these universal items, through the process of community engagement and review of the SHARE transcripts described above, the VIEW team has worked to assure conceptual and linguistic consistency with BHI items. The greater level of scrutiny via the cognitive evaluation of universal items will also allow confidence that items are capturing what they are intended to capture.

The first cognitive evaluation approach is evaluation of item structure and format as well as item order. The online electronic panel (e-panel) provides access to a large sample of female respondents which is necessary to evaluate distributional differences between both item structure of item versions as well as item order. The Bladder Health e-panel evaluates a wide range of issues associated with candidate items in the BHI, including:

- Response formation (exceptions, recall accuracy and deconstruction)
- Item preference (comparison of words and phrases, e.g., frustration v. bothersomeness)
- Response categories and impact of range of options
- Question order and evaluation of location within the instrument
- Item preference

The e-panel version includes up to 13 "experiments" whereby respondents are randomly assigned to one of 2 or 3 formats of an item. This evaluation is focused primarily on the universal items, e.g., items that every BHI respondent will be asked, including items about awareness of bladder function and perception of bladder health. These items correspond to Parts B-F of the Bladder Health Instrument. This cognitive evaluation approach uses an on-line electronic sample of respondents for data collection. This is a proprietary panel developed and managed by Survey Sampling International (SSI). The e-panel sample is a stratified random sample, with 16 strata: age (18-25, 26-44, 45-64, 65+), geography (rural, urban), and education (≤high school, >high school) and a total of 2000 interviews completed.

The second approach to cognitive evaluation of items uses face-to-face cognitive interviews (CI). The data from CI will inform refinement of item language to be consistent with women's use of terminology and phrasing, and to make sure that items are interpreted by women as they were intended to be interpreted and that the response categories align appropriately to item stems to capture relevant time frames, quantities etc. CI will address the ability of both

the symptom specific items (part G of Bladder Health Instrument) and a more refined evaluation of the universal items following any revisions made following the e-panel evaluation.

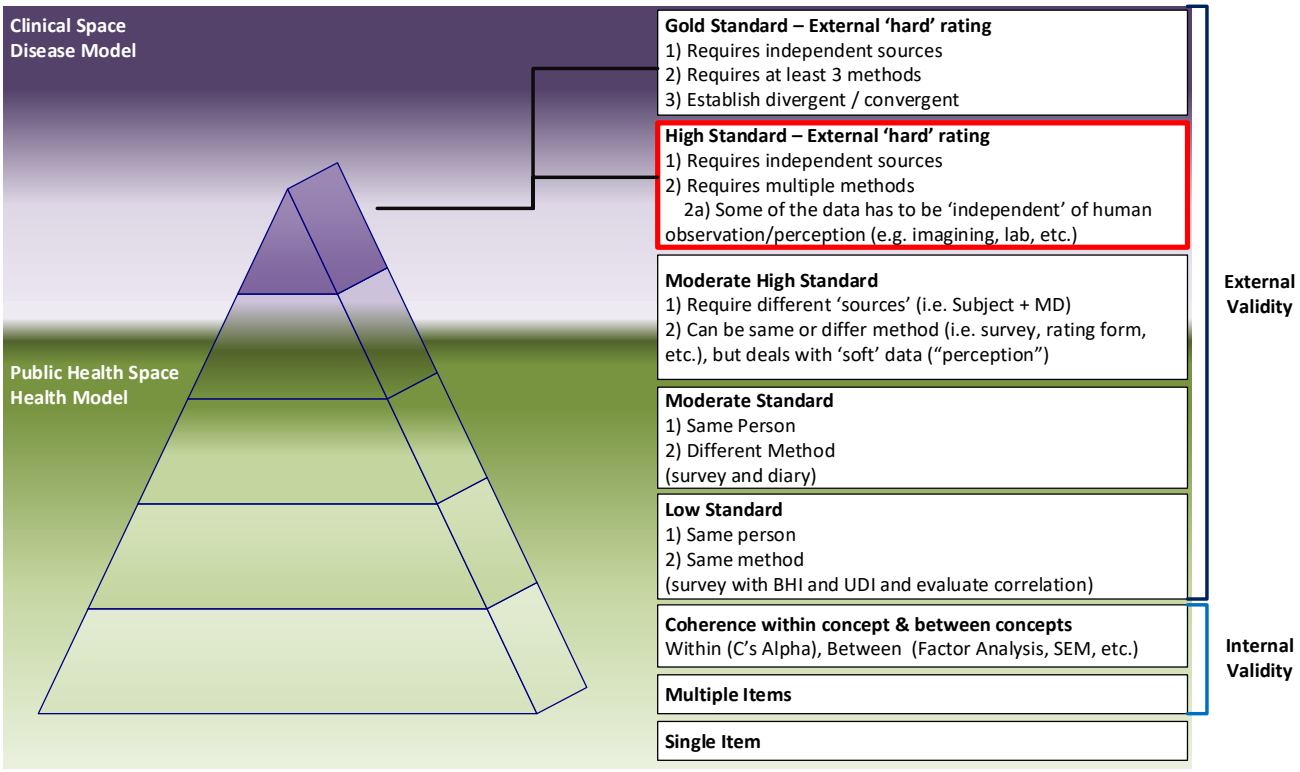
IV. Methods

IV. A. Validation approaches

In PLUS, we are seeking to ultimately create a Bladder Health Scale (BHS) to draw valid inferences about a distribution of bladder health in women and adolescent females and establish some basic inferential validity. We will use a "unified" approach to validity, which emphasizes both psychometric evaluation as well as consideration and evaluation of the usefulness, relevance and intended use of the BHS.[10,11] Any "validation" efforts may only support the assumption that the instrument is a valid measure for which to draw inference in a population that is comparable to the populations used to validate the measure[12,13] Therefore, it is essential that the populations represented in the validation samples have some equivalence to the target populations the BHI is intended to be used in.

A further consideration in validation approach is the intended use of the measure. The intended use dictates the level of evidence required for inferential validity of a measure.[14–16] For instance, a measure that is intended to be used as a population screen to discriminate between those considered at risk versus those not at risk is a standard need of public health. For valid measurement, the use of a measure to screen for potential risk requires a less stringent level of evidence than would a measure intended for use as a clinical diagnostic tool. Diagnostic tools require more rigorous levels of evidence for valid inference and use in the intended population, (e.g., ability to discriminate between presence and absence of actual disease state) than does a public health model. Figure 3 illustrates the hierarchical "levels" of evidence and associated methods for measurement validation. As noted in the figure, more rigorous methodological approaches to validation are needed when a higher level of evidence is needed. The level of evidence needed, based on intended use of measure, is indicated by levels at the lower base of the pyramid (e.g., the public health space), up towards to the tip of the pyramid with a clinical disease model of measurement (e.g., diagnostic space).

Figure 3. Levels of Evidence for Inferential Validity



The PLUS Consortium has agreed upon the need to ensure the BHS meets a high level of evidence for measurement that is required for inferential validity. While a diagnostic standard or evidence standard is not achievable with the current state of science relative to women's bladder health, validation to the high standard of evidence indicated by the red box in the pyramid, will allow valid inference to be made by BHS score in both population as well as clinical research.

Generally, criterion validity is based on evaluation of the correlation of the measure, or its constituent constructs, to an external measure. Evaluative criteria are correlation coefficients between individual BHI domains and data external to the BHI, with the specific correlation statistic being determined by the variable types. Method refers to a measurement procedure which includes consideration of both data source as well as data collection method. The referent data source for this validation is the BHI respondent self-report and the referent method is a self-administered survey. Utilization of the Multitrait-Multimethod matrix (MTMM) allows evaluation of correlation within (monomethod) and between (heteromethod) methods and sources that are related to within (convergent) and between (divergent) dimensions or traits.[17]

Subjective or "softer" levels of evidence are correlation of the intended measure (the BHI) to other measures derived from the same source and using the same method of data collection, such as the correlation of the symptom specific construct of the BHI with a validated measure of symptom severity measure based on respondent self-report. Higher levels of evidence are

required for higher standards of criterion validity of the BHI i.e., a higher level of evidence requires correlation of the BHI with multiple external and independent sources, as well as objective, independent measures, such as a quantified standing (provocative) paper towel test for stress incontinence, post-void residual and uroflow parameters. The needs for correlation of BHI scores with external and independent measures and methods requires having access to these clinical measures for BHI respondents. Study design, including sampling and recruitment (clinical evaluation sample), is dictated by this need.

The PLUS BHS is intended to measure bladder health in *both* the general population and clinical research. Additionally, the validated BHS is intended to be used in a self-administered via paper and pencil (PAPI) mode and an electronical/computer assisted (CASI) mode. As such, this study aims to validate the BHI using samples from two distinct recruitment populations (general population sample and local clinical research center sample) and assessing the two modes of administration within each.

The material below describes the rationale for the inclusion of each of the sample frames for the BHI validation study as well as recruitment/enrollment approaches for each sample frame. The general population sample will be administered and managed by the SDCC.

IV.B General Population Sample

IV.B.1. Sampling

The BHS is intended for use in the general US population and it is critical that the BHI validation include individuals who meet these criteria. At the current time, the best sampling frame available for the US general population is the US Postal delivery sequence file (DSF); it has been demonstrated to have the smallest amount of coverage error. This is an address based sample frame, which excludes nursing homes and other institutional or long term care facilities. Additionally, it is more efficient and cost-effective than alternative general population sampling methods, such as random digit dialing (RDD). This population will serve as the backbone for the psychometric evaluation; it is the sampling method that best characterizes the general US population.

To ensure the BHI is valid for use in both paper and pencil as well as electronic mode of administration, we plan to randomize participants to a paper and pencil instrument (PAPI) mailed version of the survey or a computer assisted survey instrument (CASI) web-based version of the survey. The sampling plan will include a simple stratification for geography. Using rural urban continuum codes (RUCC) to identify the general urban-to-rural characteristics for every zip code in the United States, we will use a simple 3 level characterization of geography: Urban, Suburban and Rural (including 'moderate city rural'). We will also use this sample frame to evaluate whether a mode effect is present that significantly affects item distribution and whether the psychometric qualities of the BHI are replicable between the two

modes. Sampled households will be randomly assigned to receive a PAPI mailed survey version or a URL for the CASI version.

In addition to the DSF, paradata with listed names of female residents where available, will be used to augment the sample frame.

IV.B.2. Recruitment & Enrollment

Briefly, the following global inclusion and exclusion criteria will determine eligible participants for the general population sample of VIEW:

Inclusion criteria:

- Community dwelling
- Age ≥18 years old
- Female sex assigned at birth
- Fluent in written and spoken English
- Able to read and provide informed consent

Exclusion criteria:

- Physical or mental condition that would prohibit self-administration of questionnaire either electronically or using paper and pencil (e.g. dementia/cognitive impairment/blindness/severe arthritis).
- Institutional living arrangement e.g. skilled nursing, long term care or rehabilitation center (these are not included in DSF)

The sampling and recruitment flow for the general population sample is shown in Figure 4. Following random assignment (RA) to PAPI or CASI mode, all households will be mailed a pre-notice letter including a tri-fold color brochure describing the study. The pre-notification mailing will include a $2.00 bill. Households randomized to CASI will also be asked to provide a valid email address using the provided postage paid return envelope. The return rate due to a bad address of 3% is expected. Upon notification of a return due to a bad address, the address will be removed from the sample.

IV.B.2.a. PAPI mode

One week following mailing of prenotification letter, households randomly assigned to a PAPI survey version will be mailed a validation version survey packet (BHI, criterion questions and demographics) with a request for the survey to be completed by the female age 18 or over with the most recent birthday. This mailing will included $10 as incentive and a stamped envelope addressed to the SDCC. Twenty two days later, households that do not respond to the initial validation survey will be mailed a 2nd validation survey packet. Twenty two days following the 2nd mailing, non-respondent households will be mailed a third validation survey, and 22 days following this mailing the 4th and final validation survey packet will be sent to non-respondent households, for a total of 5 contact points. It is reasonable to expect response rates to survey

mailings of 10%, 7%, 5% and 3% for each successively mailed survey, respectively, thus the number of households sampled is estimated to be 1500 in order to achieve 333 responses.

IV.B.2.b. CASI mode

Households randomly assigned to the CASI mode will be asked to provide an email address. Respondents who do not provide an email will be assigned to the "push to web" group. The push to web group will be mailed a letter that includes a URL for survey completion. Households randomly assigned to the CASI mode who do provide an email address will be further randomized to the push to web group or to receive an email with a link to the survey (referred to as the direct email group). This second randomization will allow a full comparison of respondents who fully elect to respond via electronic mode to those who do not. A total of four follow-up letters with a URL to an electronic version of the survey packet will be mailed to non-respondent households of this push-to-web group at 16 day intervals, and between the 2nd and 3rd letter sent, non-respondents will be mailed a paper version of the survey, for a total of 6 contact points. Respondents who provide an email address and who are randomized to the direct email group will be sent an email that contains a URL link to an electronic version of the survey packet. A total of four follow-up emails with a URL link to an electronic version of the survey packet will be emailed to non-respondents at one-week intervals, also with a paper version of the survey mailed between the 2nd URL and 3rd URL, for a total of 6 contact points.

Response rates are expected to differ from PAPI mode, with an expectation of a 10% overall response rate to the request for CASI completion.[15] With the use of a listed DSF frame, households will be randomized to account for expected differential response rates between paper and electronic completion. Within budget allowances, households assigned to an electronic version that do not complete the survey electronically will then be sent a paper version of the survey.

IV.B.2.c. Data Sharing with NIDDK Repository

For both the PAPI and CASI groups, participants will be randomized 1:1 to the placement/timing of receiving text regarding data sharing with NIDDK repository. One group will receive the text immediately at the end of the BHI and the other group will receive the text separately after all study surveys are completed.

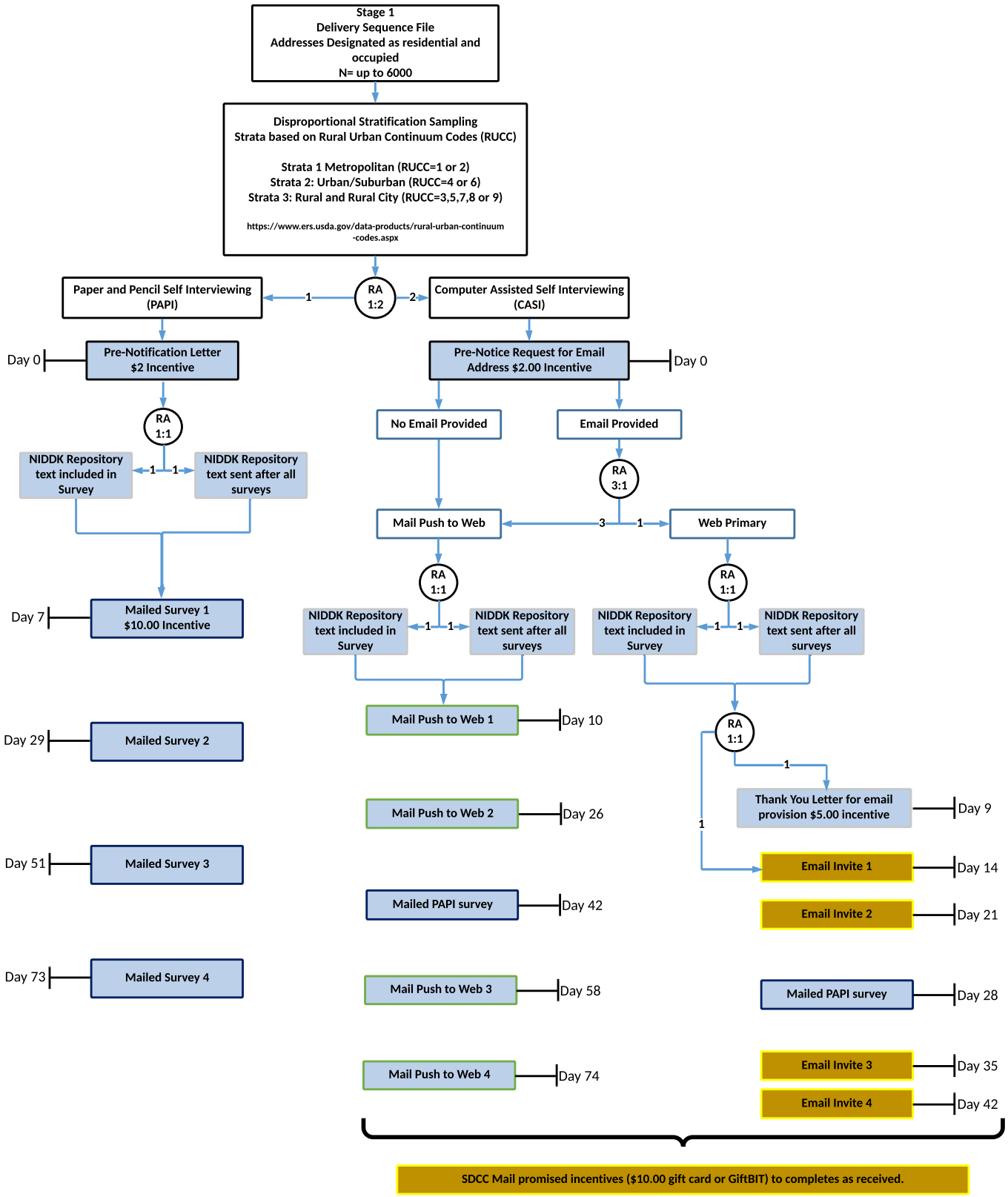


Figure 4a Stage 1 General Population Sample Plan

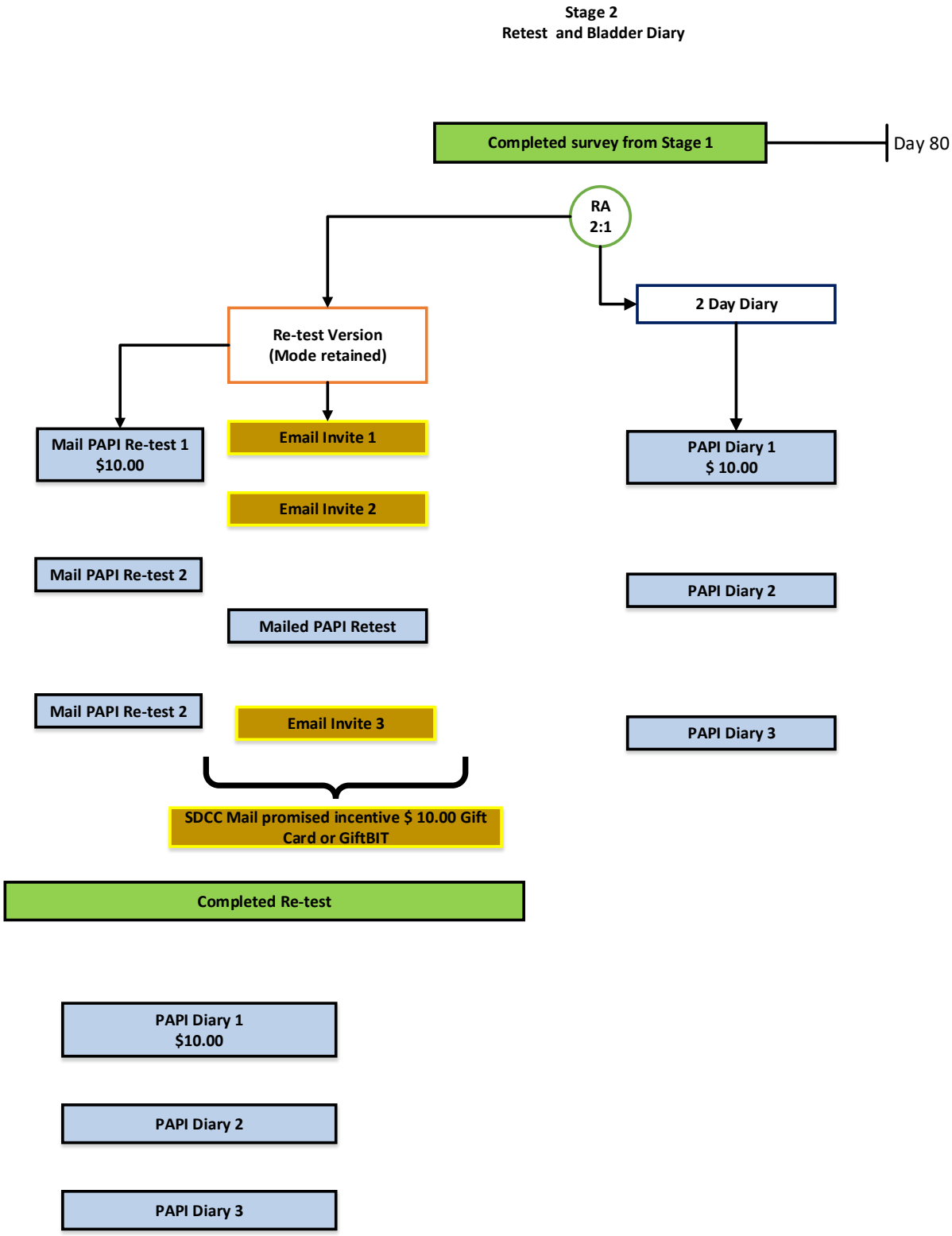


Figure 4b Stage 2 General Popolution Retest and Diary Sample Plan

IV.B.3. Retest and 2- day Bladder

All sampled individuals will be asked for permission to be re-contacted after completion of the survey. Consenting respondents who have fully completed the initial validation survey will randomly assigned to either the retest and bladder diary sample or to a bladder diary only sample. All respondents will be re-contacted within 2-4 weeks of initial survey completion. Respondents assigned to the retest + bladder diary sample will be asked to complete a re-test version of the BHI.. The re-test version will be provided in the same format the respondent used for completion of the initial validation survey (PAPI or CASI). This re-test survey will be a shorter survey than the initial validation survey as it will only include BHI items: demographic and external criterion measures will not be included in the re-test survey. The retest version will also include 2 Guyatt transition rating items which will serve as an anchor for stability in the retest reliability analysis where only stable respondents will be included. Additionally, respondents who complete the retest version of the BHI will be asked to complete a 2-day Bladder Health Symptom diary including voiding frequency, continence, sensation and voiding symptoms over 48 hours. A paper version of the 2-day Bladder Health Symptom diary will be sent to all respondents, independent of mode of completion of initial survey. Respondents who are randomly assigned to the bladder diary only will be asked to complete the bladder diary within 2-4 weeks of completion of the initial validation version of the BHI. As indicated in Figure 4b, up to 4 contact attempts will be made for completion of the retest version of the BHI and up to 3 contact attempts for completion of the bladder diary. Data from the 2-day Bladder Health Symptom diary will provide a higher level of evidence for criterion validity. Respondents will receive compensation for completing the re-test at $10 as well as the 2-day Bladder Health Symptom diary at $10.

IV.C Clinical Evaluation Sample

The other intended use of the BHS is to draw inferences related to bladder health for use in clinical research. The higher level of evidence needed for the BHI to be considered valid for inference in a clinical sample will be data collected from a sample composed of women recruited from: 1) local clinical research center communities, 2) through community partners and/or 3) undergoing clinical care for non-LUTS, LUTS and obstetric conditions. The clinical population sample is so termed due to the in-person clinical data collection associated with these participants in the validation process.

The goals of this sampling are to develop a continuous measure of bladder health, with values ranging between very healthy to very unhealthy. The following are the inclusion and exclusion criteria for the participants recruited from the clinical research centers. Of note, postpartum eligibility is independent of mode of delivery (spontaneous vaginal, operative vaginal, and caesarian deliveries).

Inclusion criteria:

- Community dwelling

- Age ≥18 years old
- Female sex assigned at birth
- Fluent in written and spoken English
- Able to read and provide informed consent
- Stand independently without human assist (e.g. cane/walker okay) for up to 3 minutes
- Get to bathroom and use toilet on own-without help from another person
- Willing to complete BHI validation survey and 2-day Bladder Health Symptom diary and 1-day Bladder Health Frequency-Volume diary prior to in-person clinical evaluation
- Available and willing to commit to an in-person evaluation within 8 weeks of enrollment
- Pregnant in 3rd trimester or recently post-partum*
- Available and willing to come for an in-person evaluation within 8-12 weeks post-partum (may be enrolled prior to delivery)*

*postpartum group only

Exclusion criteria

- Physical or mental condition that would prohibit self-administration of questionnaire either electronically or using paper and pencil (e.g. dementia/cognitive impairment/blindness/severe arthritis)
- Institutional living arrangement (e.g. skilled nursing, long term care or rehabilitation center)
- Pregnant at the time of data collection
- Diagnosis or history of bladder cancer, kidney transplant, pelvic radiation, or currently getting dialysis
- Unable to stand and toilet independently
- Current participation in a research study about bladder

IV.C.1. Clinical Population Sampling – Research Centers only

IV.C.1.a. Community population

The PLUS research centers in the consortium involved in participant recruitment and evaluation of the clinical sample are:

- Loyola University Chicago
- University of Alabama at Birmingham
- University of California San Diego
- University of Michigan
- University of Pennsylvania
- Washington University

- Yale University

Each of these sites have the capacity to recruit participants to their local site for this portion of the study. The sample will be primarily recruited from the community, however may also include recruitment from medical practices/clinics as necessary to capture a range of participants with and without LUTS. Six research centers have unique access to post-partum women due to the nature of the consortium member's practice or recruitment access. These centers will enroll the post-partum population (UAB, San Diego, Michigan, U Penn, Washington U and Yale). Clinical data will be collected from all sample participants.

Recruitment will target 3 groups: 1) women who are asymptomatic of LUTS 2) women who are symptomatic of LUTS and 3) women in the postpartum period. The intent for this targeted recruitment is for a distribution of bladder health to be represented in the validation sample from very healthy to very unhealthy. Access to a clinical sample will allow data collection for assessment of aspects of bladder health from 2 sources external to the BHI respondents as well as an independent "objective" hard rating (see Figure 3 and Table 2). These data sources are essential to data collection, providing the higher level of evidence required for the level of validation agreed upon by the consortium. While it is known that self-perception can substantially deviate from clinical impression and anatomical/physiological testing, it is a goal of the PLUS consortium to bridge the transition of a measure that is both 'valid' for public health/general population research and valid for clinical research. This requires inclusion of data sources and methods external to the self-reported BHI data collected using survey methods.

The clinical data will allow us to evaluate the ability of the BHI to differentiate between fine gradations of bladder health. An anticipated issue is the distribution of BHI scores across levels of bladder health. It is expected there will be substantial floor and/or ceiling effects in women whose bladder is unhealthy. This is data we expect to have from sampling of the general population. Normally, items with very skewed distributions are excluded from consideration in scale development as they indicate a limited ability to discern variance, although in this study we actually expect to see this skew. Evaluation of the distributional properties of items from a symptomatic sample that are expected to range across the mild to severe bladder symptoms will allow us to identify potential cut-points of items where the floor/ceiling effect disappears as bladder health changes.

Sample characteristics will be monitored sequentially by the SDCC to ensure appropriate diversity. Recruitment of non-pregnant women will use a systematic approach; when a strata quota is full (i.e., severe), individuals will be screened out and recruitment will focus on the open strata. An approximately equal distribution of women across crude degrees of bladder health will be attempted (Table 1).

Table 1. Clinical Community Sample Strata

	VIEW Site Specific Target Enrollment

	Self-report of Bladder problems				
Self-report of LUTS	Healthy	Mild	Moderate	Severe	Total
Frequency	94	15	15	15	45+
Incontinence		15	15	15	45+
Urgency		15	15	15	45+
Pain/Discomfort		15	15	15	45+
Peeing/Flow		15	15	15	45+
UTI		15	15	15	45+
Site Total	94	90	90	90	

Recent secondary analyses of the BACH database by the PLUS consortium has identified that of the 2,697 women with complete data, 16.9% reported no LUTS or interference. The large percentage of women within that community with some level of symptoms and/or interference highlights the fact that recruitment from local communities may fulfill the numbers needed to capture the distribution of bladder health, without need to recruit from specialty LUTS clinics. We anticipate that local site recruitment will initially highlight "healthy" women. If insufficient numbers of women with LUTS symptoms or interference are available, then community based recruitment of women with "bladder problems" will be implemented, followed by recruitment from LUTS specialty clinics of women seeking care for LUTS.

1). Asymptomatic: The first focal group in the clinical evaluation sample is women who, based on self-report, have no bladder problems. Each research center site will enroll women recruited from the local site using site specific recruitment strategies targeting women who meet the screening criteria. These women will be categorized as "healthy" solely on self-report of bladder problems and independent of self-report of LUTS. Therefore, women who are categorized as healthy based on self-report of no bladder problems will not need to be stratified across LUTS categories (rows in Table 1); the target is to recruit a total of 94 women who self-report no bladder problems.

2). Symptomatic: The second focal group in the clinical evaluation sample is women who, based on self-report, experience mild to severe bladder problems and are symptomatic of LUTS (urinary urgency, incontinence, frequency, voiding dysfunction, pain, or frequent UTI). The PLUS symptom screening form will be utilized to estimate mild, moderate and severe LUTS and women will be recruited using site specific recruitment strategies targeting symptomatic women either from the community/primary care clinics and/or from specialty clinics as needed in order to complete the cells in Table 1. Additionally, minimum numbers of participants for each of the 4 age cetegories will be targeted to assure adequate cellsizes for analysis.

IV.C.1.b. Postpartum population

Target sampling of a postpartum population is necessary because it is unlikely that this group would be found in sufficient number in the general population sample relative to the reference group who are considered post-partum; the US birth rate in 2016 was 12.2 births per 1000 women aged 14-44.[18] Antepartum and post-partum women are a focal population identified for

future research using a validated BHI due to the known high risk of developing LUTS in the peripartum period and evidence for successful LUTS prevention strategies. Recalling peripartum LUTS is not accurate when done at a time remote from the delivery and is problematic in that the potential for recall bias is very high, resulting in significant measurement error.[19] Adequate representation of this important population is critical to establish confidence that the BHI score will allow valid inference to be made by future studies of this population. Therefore, the sampling for the validation study includes a focal group of women who are in the postpartum period, defined for this study as 6-12 weeks post-partum, during the time of survey completion. Recruitment from the obstetric population (pregnant/post-partum) may occur simultaneously with those recruited from community and specialty clinics. The six research centers with access to these women will enroll participants. Due to the limited age range and limited range of bladder health in this population, we will use a convenience sampling technique without specific targets for distribution across LUTS. Broad demographic and racial representation will be emphasized and tracked.

IV.C.2 Recruitment and enrollment –Clinical evaluation sample

IV.C.2.a Community population

Prospective methods will be used for recruitment and data collection. Potential participants will be identified through local clinical research center practices, recruitment flyers and community engagement activities. The PLUS research centers may also partner with a family practice clinic, an OB clinic, or community centers. Enrollment will be targeted initially to healthy women (e.g., scheduling an appointment for an annual well-being visit or attending a local health fair). Other recruitment strategies may be proposed and individualized by site to maximally cover the spectrum of bladder health desired for this study. The distribution of LUTS within enrolled participants will be monitored by the SDCC and recruitment adjusted as necessary to ensure minimum representation of each symptom in the final clinical sample. Recruitment from local site specialty clinics may be necessary to fulfill the LUTS categories and will require a partial waiver of consent for pre-screening purposes to identify potential study participants with specific LUTS symptoms/severity.

Possible VIEW Recruitment Strategies:

Community and Asymptomatic Participants: letters, flyers, word of mouth, contacting women interested in prior Consortium studies SHARE and CLEAR, but who were not able to participate, university/institutional websites, institutional research centers/recruiters, social media, community partners & advocates, local community centers, flyers at community events, Craigs List, Facebook

Symptomatic Participants: direct recruitment (screening schedules) from clinical offices (Int med, family med, geriatrics, GYN/Urogyn, urology), flyers in clinical offices, institutional/clinical databases, university/institutional websites, letters, flyers, word of mouth targeting women with bladder symptoms.

Once participants are recruited and enrolled, only those who provide contact information for both home address *and* email will be eligible for randomization to PAPI vs. CASI survey administration. As participants are enrolled, they will be assigned (via block randomization by the SDCC) to either a PAPI or a CASI based mode of the BHI survey. Those who do not have email capability will be included in the PAPI arm.

Participants assigned to PAPI mode will be mailed the survey packet from the SDCC with a brief description of the components of the study, the BHI and criterion questionnaires. The survey packet will include a postage-paid return envelope addressed to the SDCC. All materials sent to participants will contain logos and contact information for the local site personnel and the PLUS consortium labeling. participants assigned to the electronic version will be emailed a unique link to online BHI completion, and contact information for the local site personnel will be available online for any questions. If the first BHI initial survey is not received by the SDCC within 15 days of mailing or one week of emailing, the PLUS site's research coordinator will contact the participant to remind her to complete the BHI initial survey and criterion questionnaires.

Upon receipt by SDCC of the completed validation survey (paper or electronic), the SDCC will notify the recruiting PLUS site's research coordinator of participants BHI completion. The research coordinator will call the participant to schedule the in-person evaluation and to mail a box containing a copy of the consent form to review along with any local instructions, the 2-day Bladder Health Symptom diary, and the 1-day Bladder health frequency-volume dairy and voiding hat along with detailed instructions for completion of all material and a local site map and parking information. Where local IRB and institutional policy allows, a $15 subject compensation for completion of the BHI will be included in the box as well. The materials will be packaged in a box and sent within one business day of the research coordinators telephone call scheduling the participant for the clinical appointment. Participants will be contacted again by the PLUS site's research coordinator 4-6 days prior to the visit to remind them to complete the Bladder Health diaries, if not done prior. Participants who have not completed the bladder diaries at the time of this reminder will be contacted again one day before their in-person appointment to be reminded to complete the diaries. Participants who have not competed the bladder diaries will be asked to reschedule their clinical appointment to a time when they will have completed the diaries.

Figure 5 depicts the flow diagram for women recruited and enrolled in the clinical research center sample. See sections IV.B.4 for details of methods of data collection. PAPI survey instruments will be returned by mail directly to the SDCC and CASI data will be captured directly by the SDCC. Tracking of returns will be real time using unique de-identified study identification numbers/bar codes and completion tracked real time via online REDCap system. Diaries will be brought into the PLUS research sites. Following informed consent, the bladder diaries will be reviewed by the Research Coordinator for legibility prior to moving forward with the evaluation procedures. The Research Coordinator will clarify with the participant any areas on the Diary that are unclear to read.

IV.C.2.b Postpartum population

Sampling and recruitment of participants for the postpartum population will occur according to local recruitment strategies. While screening assessments for LUTS will be captured, we will not attempt to recruit into particular cells based on symptom distribution. Additionally, while attempts will be made to recruit an ethnic and racially, socially and demographically diverse population of women, we will not assign minimum numbers for these categories. Monitoring of the composition of the population will be performed by the SDCC. Should significant gaps in particular populations be identified by the SDCC, local centers will modify recruitment strategies to fill the gaps.

Enrollment and randomization to PAPI or CASI survey administration will proceed in a fashion similar to the community sample. These participants may be recruited prior to delivery but will be mailed/emailed the survey packet to complete no sooner than 6 weeks post-partum. After participant completion of the BHI the in person clinical evaluation will be scheduled to take place by or before 12 weeks post-partum.

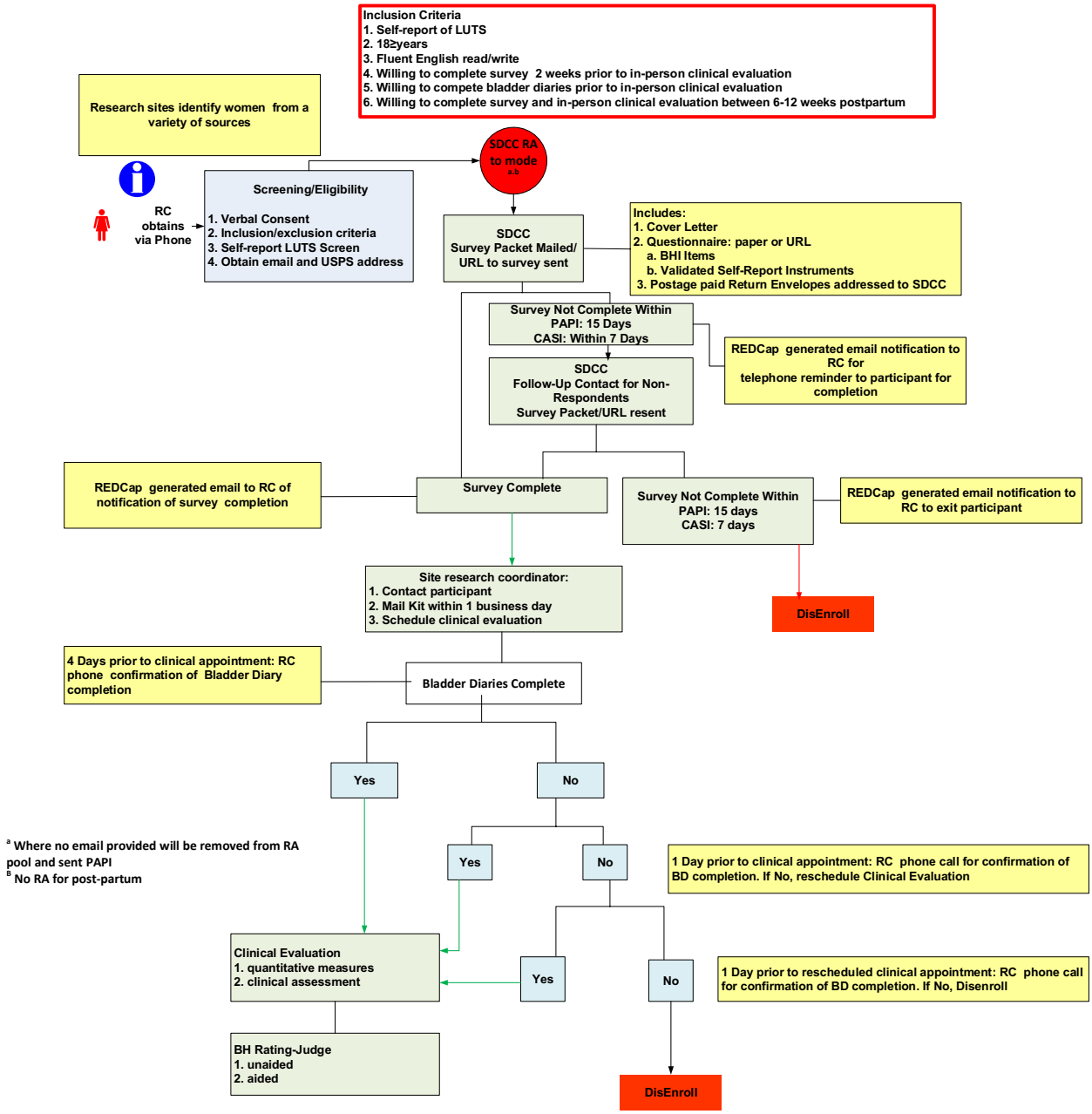


Figure 5 Clinical Sample Recruitment and Enrollment

IV.D. Measures

IV.D.1 Bladder Health Instrument

Multi-item measurement is used to assess the range of bladder health dimensions as defined by the model of bladder health adopted by the PLUS consortium. It is anticipated that the self-

administered BHI instrument (PAPI or CASI) will consist of approximately 85 items: 53-67 are items all respondents will be asked to answer, and 48 are asked only of women who self-identify as experiencing a specific LUTS. The items cover the range of bladder health dimensions specified by our model.

IV.D.2 Demographic and general medical history

Standard demographic items will be included in the survey packet mailed or emailed to women enrolled in the study, including those to assess age, race/ethnicity, and socioeconomic status. Questions on medical history covering sleep apnea, diabetes, smoking, hypertension, other pelvic floor disorders (bowel leakage, pelvic prolapse), UTI diagnosis, recurrent UTI diagnosis, antibiotic treatment, prior treatment or surgery for LUTS, will also be included in the material.

IV.D.3 External Criterion self-reported measures

While numerous validated condition specific LUTS questionnaires exist for assessing outcomes of LUTS interventions, none fully address the spectrum of bladder health in a non-clinical population. The PLUS consortium reviewed available instruments and assessed each one's ability to assess the bladder functions defined by the PLUS consortium. Validated measures selected to use as external criterion include general health items from the Medical Outcome Study (MOS) to evaluate association with the general bladder related health dimension of general bladder health.[20] Naughton et al. determined that the King's Health Questionnaire (KHQ)[21] and UDI represented Grade A/highly recommended instruments for use in LUTS research.[22] The consortium determined that the KHQ and UDI may have insufficient items for evaluation of voiding function, thus the Bristol Female Lower Urinary Tract Symptoms (BFLUTS 3 voiding items (V1-V3))[23] was deemed important to include as an external criterion measure. The PFDI-20[24] will also be included as an external criteria with expectations the UDI-6 will strongly correlate with several dimensions of bladder health and the CRADI and POPDI subscales will serve to evaluate discriminate validity.

IV.D.4.a 2-day Bladder Health Symptom Diary

The 2- day Bladder Health Symptom Diary, also based on participant self-report, is an expanded version of a voiding record assessing storage symptoms (frequency, continence, sensation of urge and pain) and emptying symptoms (initiation, flow, efficacy, sensation of urge relief and pain) along with fluid intake and absorbent product usage over two days. This data will be collected from the general population sample, and the clinical evaluation sample.

IV.D.4.b 1-day Bladder Health Frequency Volume Diary

The 1 day Bladder Health Frequency Volume Diary will assess self-measured voiding frequency and volumes along with symptoms of incontinence and absorbent product usage on a one day/24 hour diary. This diary will be mailed from the PLUS research site in a box containing detailed information about the study, a local site map and parking information, a copy of the informed consent, and a 1000 mL voiding hat with instructions for accurate measurement. The Bladder Health Frequency Volume Diary will not be completed by the general population

sample due to the low benefit to cost ratio of this measure in a general population where an external, independent clinical evaluation to interpret volume measure is not possible.

IV.D.5. Clinical tests

Clinical tests will be carried out by PLUS research sitedesignated staff. Designated staff will be required to complete test adminstration training designed by PLUS clinical investigators.

- Height and weight measurement
- Quantified standing (provocative) paper towel test (PTT) for stress incontinence[25]
- Non-instrumented seated uroflowmetry with a comfortably full bladder (min 150 mL) including print out of voided volume, urine flow time, and maximal and average urine flow rates.
- Non-invasive pre and post-void residual bladder scan
- Urine analysis (a.k.a. dipstick): pH, specific gravity, blood, glucose, protein, leukocyte and nitrites

IV.D.5.A Paper Towel Test (PTT)

The PTT will be conducted as part of the clinical tests to assess for stress incontinence. The PTT quantifies the amount of leakage as a continuous level variable. By using an ordinary standardized trifolded brown paper towel, even a small volume of urine loss down to a fraction of a drop spreads into a readily observable and measurable wetted area. All PTT will be conducted with a consistent brand of paper towels provided for this study.

Procedure:

- The clinical evaluator will need to be present to ensure proper completion of the PTT and to collect and measure the area on the towel within 10 seconds of completing the test.
- The test will be performed with a comfortably full bladder confirmed by the participant's sensation of fullness and by a bladder volume of $\geq$150 mls indicated on the BladderScan ultrasound
- Wash/sanitize hands and put on gloves.
- Place a blue pad (chux) on the floor for participant to stand on.
- Ask the participant to stand on the blue pad, with feet shoulder width apart.
- With the folded paper towel flat in your palm, ask the participant to place the paper towel lightly against the perineum (advise that the test is being done to demonstrate leakage and hence don't press on the perineum to hold back leakage).
- Be sure the participant keeps the trifold paper towel flat in their palm, and does not fold it over onto itself to create a thicker towel.

- Instruct the participant to do 3 single hard coughs.
- With gloved hands, take the paper towel from the participant and lay it on the counter for measurement within 10 seconds of the test.
- Observe and mark the paper towel with a determination of results as follows:
    - Observe the towel carefully! Sometimes vaginal secretions are transferred to the paper towel. These are noticeable as a shiny substance sitting on the surface of the towel and poorly absorbed into the towel. It is readily distinct from the dark wetted area of urine. These towels should be marked with a large **X** on the towel for no urine leakage.
    - If the towel is unmarred with vaginal secretions or with wetted area from urine, these towels should be marked with a large **X** on the towel for no urine leakage.
    - If the towel is wet from urine loss, at 10 SECONDS outline the wetted area using a ball-point pen to trace around it. If there is more than one wetted area (sometimes over the 3 coughs leakage hits a different spot on the towel), outline each of the wetted areas. Ignore that the wetted area will continue to spread beyond the marked area after 10 seconds in cases of higher volume leakage. By protocol, wetted area is determined at 10 seconds or less.
    - Using a clear plastic ruler, measure the longest and widest diameter of the circled wetted area, in millimeters (mms). If there are additional circled wetted areas, only measure and record the circled area longest and widest diameter
        i. Do not attempt to calculate wetted area by hand. Rather, record separately the length of the wetted area and the width of the wetted area on a data collection form.
        ii. Note the measurements on the Clinical Test Form under **#2 Paper Towel Test:**
            - Mark ☐ Overflow if there is any circled wetted area that ran off the edge of the paper towel
            - Length in mm: _____
            - Width in mm:_____
            - If the Paper Towel Test is not completed, check ☐ Not done and write an explanation in the space provided

    2. After the PTT is completed, the participant should be offered towels for cleaning and/or patting dry, as needed, and a chance to wash her hands.

IV.D.5.B Uroflowmetry

The Uroflow test will be conducted as part of the clinical tests to measure the volume of urine voided, the speed with which it is voided, and how long the voiding takes. The Flowstar is a stand-alone uroflow system that consists of the following equipment: 1) flow sensor, 2) flow stand, 3) funnel and urine container, 4) commode, and 5) AC adapter –power supply. Flowstar has a built-in printer connected to the Flowmeter. The sensor measures flow and volume. The test begins automatically when a urine flow (participant starts to pee) is detected. Flow and urine volume are recorded in real-time, so the results are immediately available.

Procedure:

- After the PTT is completed, the participant will be asked to urinate on the uroflow commode. The recommended voiding position is sitting.
- Immediately before the test is conducted, the participant should be asked to wipe off any vaginal secretions with a dampened towelette used for collecting clean catch urine specimens.
- Provide the following instructions before leaving the room for the participant to void in private:
    - Wipe urethra and vaginal area clean as if they were collecting a urine specimen at the doctor's office
    - Sit on the commode
    - Void completely like you normally would
    - Discard tissues where appropriate for the setting
    - Come back to exam room (if uroflow in separate room)
    - Please do not touch any of the uroflow equipment
- Take the report from the printer. If extra copies of the results and curves are desired, press the Print button. On the printout, the flow curve and voided volume curve are printed in real-time. The voided volume is the total volume voided.
- Note the following parameters on the Clinical Test Form under **#3 Uroflow**:
    - Check Box for: Printout ☐ Yes ☐ No
    - Voiding Time: _____ sec
    - Flow Time:_____sec
    - Time to Peak Flow :_____sec
    - Peak Flow: _____mL/sec
    - Average Flow: _____mL/sec
    - Voided Volume:_____mL
    - If the Uroflow is not completed, check ☐ Not done and write an explanation in the space provided

- Save at least 30 mL of urine from the container to test for Urine dipstick (see procedure).
- Tape the report on the Uroflow Printout Form (Form 4B). After the In-person Visit, copy and place this form in participant's chart.

IV.D.5.C Bladder Volume Assessment (e.g Bladder scan)

The bladderscan will be conducted as part of the clinical tests to assess for the adequate urine volume (≥150 mL) to perform the Paper Towel Test (PTT) and to determine post-void residual (PVR) following a uroflow test. BladderScan Prime Plus, a 3-D ultrasound system, will be used to measure the amount of urine in the bladder. The core components of the system are a console with a touch-screen display, a scanning probe containing the ultrasound transducer, a printer and a battery charger with rechargeable lithium ion batteries. Before the visit begins, turn on Scanner and check battery life. Go to settings and note accurate date, select B mode and ensure that print settings are set to B mode. Check printing paper for adequate amount to complete scheduled visits for day.

The participant will self-report bladder fullness or urge sensation (inferred by participant's level of need to void). In order to complete the Paper Towel Test, there should be ≥150 milliliters (mL) of urine in the bladder. If there is < 150 mL in the bladder, then the participant is instructed to drink at least 8 ounces of fluids (e.g. water, juice, etc.). The bladder assessment is repeated after allowing for time for filling of the participant's bladder. Generally, the bladder fills at 30 to 100 mL per hour. Depending on how low the volume is before starting the scanning, the participant may need to wait an hour or more for a full enough bladder. The bladder scan may be repeated approximately every 15 to 30 minutes until there is $\underline{\geq}$ 150 mL of urine. The participant can continue to drink fluids until there is $\underline{\geq}$ 150 mL of urine.

Procedure:

- The participant is instructed to disrobe from the waist down and a sheet provided to cover below the pubic bone. The participant may keep on socks during the procedure if desired.
- Wash/sanitize hands and put on gloves.
- Turn on the scanner and select from which side of the participant you are scanning**.** Ensure the diagram on the hand held device is oriented in the same direction as the participant.
- With the participant lying in the supine position (head elevated no more than > 30 degrees), with abdominal muscles relaxed, the participant's pubic bone is palpated.

- An ample quantity of ultrasonic gel should be placed with as few air bubbles as possible midline on the participant's abdomen approximately 3 cm (one inch) above the pubic bone. Gel can alternately be placed on the probe itself (or both abdomen and probe).
    - Be sure there are no areas of gaps between the probe and the participant's skin, and that enough pressure is applied to maintain adequate skin contact until the scan is completed. If necessary, more gel should be added to insure proper contact.
- The probe is held by grasping with the cable running up the wrist and forearm.
- The probe is gently pressed to the lower abdomen through the gel just above the pubic bone aiming away from the participants head and more toward the feet and the diagram on the top of the scanning probe should be oriented in the same direction as the participant. The probe cable should be oriented at 90 degrees to the digital plane of the participant
    - If an obese (large girth) participant is being scanned, lift as much abdominal adipose tissue out of the way of the scanning probe as possible. More pressure should be applied to the probe in order to reduce the amount of adipose tissue through which the ultrasound probe is passed.
- The green button on side of probe is then pressed or press the "Scan" icon on the console screen. The real B-mode ultrasound image appears on the console screen. Target the bladder by doing the following:
    - Angle the probe slightly from the participant's left to right until the dark bladder area is centered in the vertical green line on the aiming screen. Once the bladder is centered, angle the probe slightly up or down the participant's midline to obtain the largest possible dark area.
    - A green outline (*BladderTraq*) will appear around the detected edges of the bladder.
    - When the largest possible dark area is obtained, the green probe button should be pressed or tap icon "Scan" on the screen to allow the scanning process to begin.
    - The probe should be held securely while the scan is in process.
    - The end scan tone sounds when the scan is completed.
    - Press the icon "Done" on the console screen.
- Aiming guiding is successful if bladder is centered in the field of view and <u>all</u> bladder edges are visible. There should be no grey areas. Bladder scanning is then successful and the results optimized for accuracy. If the bladder volume is outside the edges and is >150 mL there is no need to repeat the scan. If the bladder volume is <150 mL and is not centered in the field, the scan should be repeated to obtain an accurate measurement.
- Press the icon "Print" on the console. Take the report from the printer. If extra copies of the results are desired, press the Print icon a second time.

- Tape the report on the Bladder Scan Printout Form (Form 4A). After the in-person Visit, copy and place this form in participant's chart.
- Note the following parameters on the Clinical Test Form under **#1 Bladder Scan volume prior to Paper Towel Test:**
    - Check Box for: Printout ☐ Yes ☐ No
    - Volume: _____mL
    - Check Box if ☐ scanned volume is < 150 mL and the participant was unable to "hold or wait for scan prior to voiding"
    - If the BladderScan is not completed, check ☐ Not done and write an explanation in the space provided

    Participants should be reminded when scheduling appointments to come in with a full bladder. If participant empties her bladder prior to the procedures, then they should be asked to drink 8 to 16 oz of fluids and rescanned when they feel full or after 15 to 60 minutes until ≥150 mL of urine.

Checking for accuracy: If inaccurate, need to repeat procedures until no error messages For the pre Paper Towel Test, if scanned volume is >150 mL, rescanning is not necessary, but if scanning is inaccurate for the post void residual test, the test should be repeated until accurate.

- Yellow "greater than" (>) symbol appears: The actual bladder volume may exceed the displays result. The participant should be re-aimed and re-scanned if this happens.
- Bladder not centered in the field of view (within green tracker): Angle the probe in the direction of the bladder on the display in order to optimize results.
- Edge scan: if one side of the bladder is not within a field of view, then a portion of the bladder was not included in the scan. The system displays a "greater than" symbol (>) before the measured result, indicating that the actual bladder volume may exceed the displayed result. The probe should be moved or angled in the direction of the bladder on the display in order to optimize results.

Pubic bone interference: if a grey area appears, this indicates that the pubic bone is inside the field of view. Although the bladder may be centered and measurement may be complete, there is the possibility that the pubic bone is obscuring part of the bladder. The system displays a "greater than" sign (>) before the measured result, indicating the actual bladder volume may exceed the displayed result. The probe may be moved or angled in order to optimize results.

IV.D.6. Clinical judgement of bladder health

All clinical evaluation sample participants will meet with a site specific bladder health rater or judge. While we previously used the terms clinician rating or expert rating, in the measurement paradigm judge is the usual term. Use of the term judge avoids the confusion associated with

the term clinician and expert. A judge is someone who has a point of view which makes their observation a viable standard which can be used as a criterion. For PLUS, a judge is a health care provider who as part of their work/practice are expected to be able to respond to basic questions or statements about the bladder from a person. Operationally, we would consider any provider (NP, PA, CNM, DNP, MD, DO) in family practice, internal medicine, geriatrics, obstetrics-gynecology and urology. Female Pelvic Medicine and Reconstructive Surgery specialists include urologists with additional fellowship training and obstetrics-gynecologists with FPMRS fellowship training. The expected heterogeneity among judges is more generalizable than results from standardizing an evaluation. We have a spectrum of judges within PLUS, and additional judges may be selected from a site as necessary. It is recommended that each judge complete a minimum of 8 ratings and a maximum of 18 ratings for the non-postpartum sample, and a minimum of 5 and maximum of 10 for the post-partum sample.

The judge will provide two initial unaided ratings of bladder health: a circumstance/context adjusted rating of bladder health followed by an absolute rating of bladder health. No specific script or checklist for this rating will be mandated; each PLUS research site and the judge may use whatever method they generally employ in order to assess their patients with LUTS. For each rating, the judge will be asked to provide the 3 most important factors that contributed to the rating. Following the completion of these ratings, the judge will generate a third and fourth rating aided by information from the 1-day Bladder Health Frequency Volume diary, clinical test results and any other data available from standard of care clinical practice evaluation. The 3rd rating will also be an adjusted rating and the 4th rating will need an absolute rating, as well as description of the 3 most important factors that contributed to each rating. The 3rd and 4th rating will preferably be on the same day as the 1st and 2nd rating but may be on a separate day but within 1 week of the evaluation visit.

Where practically possible, the participant-judge interaction will be scheduled to and actually occur prior to the administration of clinical tests. This is needed to minimize potential for test effects to influence participants interaction wth the judge. Further, if the tests are performed before the clinical interview, the results may not be provided to the judge until after have they have completed the interview and done the initial 1st and 2nd ratings.

Participants recruited from community or from non LUTS practices who report bothersome symptoms and request evaluation and treatment will be referred for clinical care per the individual PLUS research site's practices. For women recruited from LUTS medical practice/clinic, who were already scheduled for a new patient visit, any additional testing and evaluation beyond the clinical tests outlined above may be performed on the same day as the research appointment and will proceed per the provider's usual clinical practice. Information

gained from routine clinical care may be used in conjunction with clinical data collected for research purposes to inform the expert evaluator's 3rd and 4th rating of bladder health.

IV.E. Analyses

All analyses will be conducted on the general population and the clinical evaluation sample separately, data from the different sources will not be pooled. The final factor structure of the BHI scale will be determined based on the general population data. While the clinical evaluation sample data will not be pooled with the general population data, the clinical evaluation sample data provides a cost-effective means of conducting some external validity tests, i.e., it is not feasible to collect data for frequency/volume diaries or clinical evaluation measures or expert evaluator ratings for the general population sample.

IV.E.1. Distribution evaluation

IV.E.1.a. Ceiling and floor effects

Items will be evaluated for ceiling and floor effects. It is expected that for asymptomatic women many items will demonstrate a pronounced ceiling effect. Therefore, item response of asymptomatic and symptomatic women will be evaluated separately. Within symptomatic respondents, it is expected that those who report mild LUTS will demonstrate greater ceiling effects than respondents who report severe LUTS. Items that exhibit greater than 85% response in the tail of the distribution within severe LUTS respondents will be removed. Comparison will be both between and within mode and between and within sample frames.

IV.E.1.b. Item missing values

Items with missing values, i.e., no response indicated, will be reviewed. A respondent's failure to provide a response to a question item can pose a threat to the scale or sub-scale (dimension or factor specific item grouping) to construct validity. Generally, statistical methods used to evaluate dimensional validity require that a respondent have complete data on all items included in the evaluation for the respondent's values to be included in the analysis. Therefore, where item nonresponse for specific items is high enough (10% or more) to impact the analysis, the nature of missing data will be evaluated. This evaluation will initially consider item location relative to questionnaire layout and design, branching and item applicability, and item sensitivity level or perceived threat to respondent. Where item nonresponse patterns are not explainable by any of these context based effects, a more formal evaluation of the nature of missingness will be considered prior to determining the scale scoring method.

IV.E.2. Reliability and Validity

IV.E.2.a. Internal Validity

Evaluation of internal validity, often referred to as reliability, will include comparison both between and within mode and between and within sample frames. The analysis for the three distinct phases of internal validity evaluation are described below.

i. Internal Consistency: Items that are expected to measure a domain or construct should be related to each other in a systematic way. Internal consistency of item groupings will be evaluated with Cronbach's alpha. If the alpha is > 0.9 for emergent factor groupings, we will evaluate the range of meaning across the different items within the grouping to identify if items that are believed to be different are, in reality, subtle variations of the same question. This will include comparison of the item total correlations and subsequently correlations between the individual items. If the alpha is < 0.4 we will proceed with the assumption that the a priori expectation of items as related is rejected. In this instance, further factor analyses will be used

to identify any distinct sub-dimensions that we did not anticipate or item groupings that do not form a coherent measure of the intended construct.

ii. Test-retest Reliability: Given no reason for change a measure will reproduce the same value that it did at a prior point in time. Tukey's HSD test with threshold alpha ≤0.05 will be used to evaluate test-retest reliability for all items. Concordance of individual responses will also be evaluated. We plan to allow a 2 week to 1 month window for test/re-test.

iii. Internal Dimensional validity: The focus of this analysis is to determine if our prior conceptualization of bladder health and the items developed for measuring bladder health, are supported across the entire pool of items. We will accomplish this identification of dimensions of bladder health, analytically indicated by factors, from our data. Factor analytic methods will be used to identify a factor structure of the BHI. While we may hypothesize what the dimensions of bladder health are, *a priori* this cannot be known. Alpha analysis identifies group of items (potential factors) having 'internal reliability', although it does not identify if the underlying relationship between items intended to measure one dimension are differentiated from items intended to measures other dimensions. Therefore, for the dimensions that meet the Alpha threshold ($\alpha \geq .40$), we will run a confirmatory factor analysis to evaluate the actual factor structures against our hypothesized dimensions. While it would be highly unusual for our theorized model to hold up to an initial unrotated factor model, we believe the work on developing a theoretical model of bladder health warrants a preliminary comparison of the model to the data. For the dimensions or factors that do not meet the minimum Alpha criteria, we will run an exploratory factor analysis to identify any systematic and meaningful relationships between groups of these items, that is, if any meaningful factors emerge. Following this we will pool all items together and conduct an exploratory factor analysis.

The sequential and iterative nature of item selection will be guided by both theoretical consideration, practical considerations and as well as statistical criteria. To reduce the risk that factor analysis can produce results that appear meaningful but may be potentially very misleading, we will have multiple investigators utilizing multiple methods (PFA/PCA) and rotation methods (orthogonal/oblique) simultaneously, working independently and periodically comparing and discussing results. Several criteria will be applied in evaluation of dimensions and factor retention: the Kaiser-Guttman rule (Eigenvalues>1.0), factor loading thresholds of 0.60/.40, scree plots, Kaiser-Meyer-Olkin (KMO) residuals off-diagonal partial correlations measure of sampling adequacy >0.70.[14,26,27] This analysis is iterative: a single item is dropped and the structure evaluated, the item is put back in and another item is dropped and the structure is evaluated and compared to the structure of the prior model. This is a time intensive and elaborate processes, but necessary to prevent drawing conclusions about factors and the items associated with them as being optimal or even meaningful, when in fact they are not.[28]

Additionally, we will repeat the analysis with the population partitioned as absence of LUTS to mild, moderate and severe presence of LUTS and re-run the models to determine whether the

factor structures and associated items are constant across various levels of symptom levels. The goal is to identify a set of items and factor structures that remain stable across groups with varying levels or LUTS, including the absence of LUTS; to identify a satisfactory final factor solution that will be inclusive relative to degree of LUTS. If systematic differences are found across degrees of LUTS we will evaluate the sources of variance item by item to arrive at a final factor solution that does not affect variance in the general population.

IV.E.2.b. Mode variability

The final validated bladder health scale is intended for use in both general population research as well as regional or clinical research. One planned use by the consortium is to establish a distribution prevalence of bladder health in a nationally representative sample of women. This requires use of a sample frame with minimal coverage error, which is only available with mailing addresses. Therefore, the validated BH scale must be able to be administered in a paper version by mail (push to web from mail contact response rates are abominably low). Alternatively, the BH scale is also intended to be used for clinical research, where an electronic version of the BH scale can be more successfully administered. As such, both validated PAPI and CASI modes of the scale are needed.

To the extent that a computer may increase respondent control or increase efficient use of completion time, this may produce some measurement error, but not threaten measurement validity. But if use of a computer (or a paper scale version) alters the distribution of BHI scores, or either mode is found to be a source of systematic variance in BHI score, validity may be threatened.[7] Response to each individual items will be compared between PAPI and CASI for mode differences. Bland-Altman plots and analysis will be used to compare mean differences using limits of agreement for 95% confidence intervals of items across PAPI and CASI mode. Items that do not fall within LoA will be flagged as being a potential problem for psychometric evaluation. Additionally, psychometric evaluation of dimensions will be done independently for each mode and the final factor structures, including Cronbach alpha, the Kaiser-Guttman rule (Eigenvalues>1.0), factor loading thresholds of 0.60/.40, and scree plots, will be compared. If the structures are the same, then it is assumed that, while mode differences exist for some items, at the aggregate level the internal relationship between items for each factor are consistent. If the structures are not similar, the focus will be on determining if the items that demonstrate significant mode effects do in fact contribute to the structure difference, accomplished by evaluating the factor structures when these items are removed, using the same threshold criteria list above. The goal is to have a single instrument that is capable or working in either PAPI or CASI, recognizing the possibility that measurement differences between modes could result in sub-scales containing different items for each mode. A comparison of responses by mode: PAPI and CASI, will be made both between and within responses from the general population sample and the clinical sample and will also include comparison of item missing values.

IV.E.2.c. External Criterion Validity

In PLUS, we are seeking to establish validity to draw a valid inference about a distribution and establish some basic inferential validity. We will operationalize this using a multitrait-multimethod matrix (MTMM) approach[17]. For validation of the BHI, each dimension is considered a trait. Method refers to a measurement procedure, which includes consideration of both data source as well as data collection method. The referent date source is the BHI respondent and the referent method is a self-administered survey.

The following are hypothesized dimensions of the PLUS model of bladder health.

General bladder function (BHI section B)

Bladder and general day-to-day life (BHI section C)

Specific Impact: Your bladder and specific activities (BHI section D)

Psycho-Social: Your bladder and mind (BHI section E)

Symptom Specific: Bladder performance (BHI section G)

Criterion measures based on respondent self-report that have been identified as expecting to correlate with the dimensions of bladder heath are included in Table 2.

Table 2. External Criterion Measures to Evaluate Association with Dimensions of BHI

BHI Section Description	Criterion	Prediction	
		Healthy	Not Healthy
B: General Bladder Health	MOS Global Health	0	X
	MOS Comparative Health	0	X
	MOS Health Transition	0	X
	Section G-LUTS Frequency	X	X
	Section G-Chronicity	X	X
	Section G-Disruption	X	X
	Section G-History	X	X
	KHQ	X	X
	PFDI-UDI	0	X
	BFLUTS V1-V3	0	X
C: Daily Life Impact	MOS Function	0	
	Section G-LUTS Frequency	X	X
	Section G-Chronicity	X	X
	Section G-Disruption	X	X
	Section G-History	X	X
	KHQ	X	x
	PFDI-CRADI, POPDI (discriminant)	0	0
	BFLUTS V1-V3	0	X
D: Specific Activity Impact	MOS Function	0	X

	Section G-LUTS Frequency	X	X
	Section G-Chronicity	X	X
	Section G-Disruption	X	X
	Section G-History	X	X
	BRFSS Physical Activity	X	X
	KHQ	X	X
	Hedonic well-being	X	X
E: Emotional	MOS Emotion	0	X
	Section G: Presence or absence of LUTS		
	Section G-LUTS Frequency	X	X
	Section G-Chronicity	X	X
	Section G-Disruption	X	X
	Section G-History	X	X
	Eudemonic well-being	X	X
Legend 0=not correlated; X=correlated MOS=Medical Outcomes Study; PFDI-20= Pelvic Floor Distress Inventory; BFLUTS=Bristol Lower Urinary Tract Symptoms: KHQ=Kings Health Questionnaire; BRFSS=Behavioral Risk Factor Surveillance Survey (2009)			

Additional external criterion measures using the referent source with a different method (tracking log) include the bladder diaries, described on page 26. Criterion measures using a source other than the referent include clinical judgement of bladder health rating, a 0-10 value, an external "soft rating" using both a different source as well as a different measure. Inclusion of this external measure meets the moderate high standard of evidence specified in the validation pyramid (Figure 3, page 13). Inclusion of "hard" measures, provided by clinical tests, described on page 27, meets the high standard of evidence specified in the pyramid. Inclusion of multiple heterogeneous criterion measures also allows population of an expanded MTMM matrix.

The matrix below (Figure 6) indicates the analytic parameters of the MTMM that will be used. This is an abridged version of the full matrix due to space constraints. The matrix column and rows correspond to each method used to evaluate each dimension of the BHI identified in the process of establishing dimensional validity. As previously mentioned, the ability to discern between trait variance versus measurement variance is contingent upon the use of more than one trait and more than one method. Additionally, data from general population and clinical evaluations samples will not be pooled. Due to group differences in sampling modalities, we will use assumptions of unbalanced design and run ANOVA.

Reliability Diagonal Monotrait-monomethod values		Validity Diagonal Monotrait-heteromethod values			Heterotrait-Monomethod Triangles			Heterotrait-Heteromethod Triangles		
Methods		Method 1 (Survey) (M1)			Method 2 (Expert Rating) (M2)			Method 3 (Uroflow) (M3)		
	Traits	General BH (T1)	Daily Life Impact (T2)	Symptom Specific (T3)	General BH (T1)	Daily Life Impact (T2)	Symptom Specific	General BH (T1)	Daily Life Impact (T2)	Symptom Specific
M1	General BH (T1)	M1T1								
	Daily Life Impact (T2)	M1T2*M1T1	M1T2							
	Symptom Specific (T3)	M1T3*M1T1	M1T3*M1T2	M1T3						
M2	General BH (T1)	M2T1*M1T1	M2T1*M1T2	M2T1*M1T3	M2T1					
	Daily Life Impact (T2)	M2T2*M1T1	M2T2*M1T2	M2T2*M1T3	M2T2*M2T1	M2T2				
	Symptom Specific	M2T3*M1T1	M2T3*M1T2	M2T3*M1T3	M2T3*M2T1	M2T3*M2T2	M2T3			
M3	General BH (T1)	M3T1*M1T1	M3T1*M1T2	M3T1*M1T3	M3T1*M2T1	M3T1*M2T2	M3T1*M2T3	M3T1		
	Daily Life Impact (T2)	M3T2*M1T1	M3T2*M1T2	M3T2*M1T3	M3T2*M2T2	M3T2*M2T2	M3T2*M2T3	M3T2*M3T1	M3T2	
	Symptom Specific	M3T3*M1T1	M3T3*M1T3	M3T3*M1T3	M3T3*M2T1	M3T3*M2T2	M3T3*M2T3	M3T3*M3T1	M3T3*M3T2	M3T3

Figure 6. The MultiTrait-MultiMethod Matrix

Our approach is based on the work of Campbell relative to the use of multiple-operationalism (multiple items to assess the range of a concept) and Campbell and Fiske relative to evaluation of validity.[17] It must be noted, in their work as with other work such as Cronbach's initial work on Alpha, many of these tests are defined as internal validity checks, which over time has come to be identified primarily as internal reliability.

The following are some general interpretations of correlation values indicated in the MTMM matrix. To note, the term 'trait' used by Campbell and Fiske is equivalent to our term 'dimension.'

- Reliability diagonal (MTMM, green): A measure is internally consistent, all the components measure the same thing.
- Heterotrait Monomethod (HTMM, blue): correlation between 2 measures of a single method but different dimensions. This is an indication of discriminant validity.
- Monotrait Heteromethod (MTHM, gold): the correlation between 2 measures of the same dimension using differing methods. A high correlation indicates convergent validity.
- Heterotrait Hetero method (HTHM, grey): correlation between different dimensions using different methods.

Generally, we expect to see the reliability diagonal demonstrate the highest correlation value and Heterotrait Heteromethod correlations to demonstrate the lowest correlation value.

To provide evidence of the High Standard of Validity, represented in Figure 3, requires satisfying convergent validity: confirmation by independent measurement methods. These validity diagonals are indicated by the monotrait-heteromethod cell values, denoted in gold. The validity diagonal value is the correlation of a dimension as measured by 2 different, independent methods. These values should be ≥0.20.

As noted in Figure 3, the level of evidence needed to achieve the highest level of evidence requires both convergent validity as well as divergent or discriminant validity. Discriminant validity requires evidence meeting the following three criteria:

1. Value of validity diagonals must be larger than the value in its row and column of the heterotrait-heteromethod triangle. This means the validity value for a dimension must be higher than the correlation between that dimension and other different dimensions measured by a different method. For example, the value of cell M2T2*M1T2 must be larger than the values in M2T2*M1T1, M2T2*M1T3, and the values in M2T2*M1T2, M2T3*M1T2
2. A dimension correlates more highly with the same dimension measured with a different method than it does with a different dimension measured with the same method. This is a comparison of the validity diagonal value for a dimension with the dimensions values in the heterotrait-monomethod triangle.
3. The same pattern of dimension intercorrelation be evident in all heterotrait triangles of the monomethod and heteromethod blocks. The monomethod blocks are comprised of the reliability diagonal and the adjacent heterotrait-monomethod triangle. The heteromethod blocks are comprised of the validity diagonal and the two heterotrait-heteromethod triangle lying to either side of it.

The nature of the validity evaluation proposed by Campbell and Fiske is complex compared to methods often used in health measurement, though we feel this level of rigor is justified. Rather than validating a disease specific measure our goal is to develop and validate a health measure in an area with little to no existing health measures exist. Psycho-social measurement often find that things are related solely because they happened to have both been measured and compared. This often leads to the identification of spurious relationships, or mistaking a mediating relationship. The MTMM matrix approach attempts to rule out these spurious or mediating relationships to the extent possible. The MTMM matrix creates conditions in which we should observe no-relationship, and if a relationship is found to exist in that condition it then becomes a flag for further evaluation. The further evaluation identifies whether 1) we have conceived of the world incorrectly and things that we thought were un-related actually are related, 2) moderating relationships, or 3) evidence that the measure does not adequately

discriminant between certain things. The MTMM matrix approach also has similar implications when measures should converge (be positively related) or diverge (be negatively related). In setting these standards, we rely not just upon pre-supposed 'positive' criterion measures, a very low bar, but the bar is raised in that the measure must demonstrate no or negative relationships with other measures, a higher level of evidence.

IV.E.3. Sample size and composition justification

The intended use of the BHS is to draw inferences related to bladder health across the **US general population** of women who are over the age of 18, as well as for use in clinical research. The **general population life-course** for this Phase of the validation has been identified by the PLUS consortium based on the following strata:

- 18-25 Emerging adult
- 26-44 Young adult and adult
- 45-64 Midlife and late adult
- 65+ Old

Sample sizes are based on several criteria. The basic rule of thumb for psychometric analysis (primarily correlation analyses) is 10 participants per survey item[29] although less conservative estimates of 5-10 participants per item[30] have been suggest. The BHI contains 2 general categories of question items: those asked of all respondents (global), and those that respondents will be branched into based on response to earlier screening type questions (LUTS-specific). The minimum number of global items under consideration for inclusion the BHI that a respondents could be asked is 53, with an additional 14 items a respondent could be branched into. Therefore the maximum number of global items that will be evaluated for inclusion in the BHI is 67. The number of global items is expected to be substantially reduced through the cognitive evaluation process. The minimum number of LUTS specific items under consideration for inclusion the BHI that a respondent could be asked is 6 (no experience of any LUTS), with up to an additional 12 items a respondent could be branched into, dependent on the number of LUTS experienced. Note it is unlikely that all of the symptom specific items will be applicable to or answerable by all respondents. Therefore, the maximum number of LUTS specific items that could be evaluated for inclusion in the BHI is 18. With a total maximum of all items included for evaluation for inclusion in the BHI being 85, the proposed sample size in aggregate, of n=1202 is sufficient under the most stringent or conservative criterion. Conducting factor analysis on each of the general population sample (n=694) and the clinical evaluation sample (n=508), still provides over 8 and 6 participants per item evaluated, respectively. This estimate is also conservative in that is assumes respondents will have experienced all 6 LUTS and therefore will be branched into the additional 12 LUTS items.

The sampled number of completes will be as follow:

Sample Completes	PAPI	CASI	Total
General Population	333	361	694
Retest	*≥75*	*≥75*	*200*
Clinical			
General	177	177	354
Post Partum			154
Total			1202

IV.F. Informed Consent

IV.F.1 General Population Surveys and Diaries

Informed consent of potential participants from the national sample frame is considered implied. The elements of consent (voluntary nature, confidentiality, risks and benefits) will be described at the introduction of the written and online modes and survey response will be considered consent. Participants will be asked during the initial survey to agree to additional contact for the re-test and the Bladder Health Symptom Diary. Participants will be provided a contact number and email to opt out at any time.

IV.F.2. Clinical Evaluation Sample

For women recruited by Research Centers from communities and non-clinical settings, waiver of written consent will be requested for completion of BHI, criterion measure surveys and diaries that occur prior to in person evaluation. For the women recruited from specialty clinics and for the post-partum sample, a partial HIPAA waiver will be requested from the IRB to access protected health information for identification of potential participants and recruitment purposes. Waiver of written consent will be requested for collection of surveys and diaries from women recruited from the LUTS specialty clinics and pregnant/post-partum population. Written consent to participate will be required at the time of presentation for in- person clinical evaluation. Consent forms will be included in the mailed material to participants, who will be asked to review them prior to the in-person visit and formal informed consent process will occur in-person where participants will have opportunity to ask questions prior to signing the consent forms and before proceeding with additional study measures. Once all questions have been answered, if the individual is interested in continued participation and has signed the consent form and HIPAA authorization for access to medical records, they will be given a copy of the consent documents along with experimental participants' bill of rights, including the contact numbers of the investigators, in the event of questions or concerns.

IV.G. Compensation

Given the balance between composition and sample size the compensation amounts are subject to revision per budget allowance. Current compensation for completion of BHI and diaries are as follows:

General population sample
- Initial Mailing $2
- E-mail received from CASI group and assigned to online completion $5
- Initial BHI survey $10
- Retest BHI survey $10
- 2 Day Bladder Health symptom diary $10

Clinical evaluation samples:
- Initial BHI and criterion measure survey completion $15
- 1 day Bladder Health frequency-volume and 2 day Bladder Health symptom diary completion $35
- In clinic evaluation $50

We anticipate some participants may not complete every study measure or visit. Thus, compensation will be prorated with a maximum of $100. The type of reimbursement will be managed by each research center and may include gift cards or checks depending on local standards.

V. Data management

The general population validation processes will be managed entirely by the SDCC including IRB submission, recruitment, survey administration, and data collection. The local clinical research center validation process will be instituted at each of the PLUS Clinical Research Centers in coordination with the SDCC who will manage central IRB submission, recruitment, study implementation and data collection. Each site may utilize different recruitment strategies for including women with and without LUTS according to local practices and needs of the consortium.

Data management in general will be facilitated by REDCap, a secure web interface for building and managing online surveys and databases with data checks used during data entry to ensure data quality. REDCap includes a complete suite of features to support HIPAA compliance, including a full audit trail, user-based privileges, and integration with the institutional LDAP server. The MySQL database and the web server will both be housed on secure servers operated by the University of Minnesota Academic Health Center's Information Systems group (AHC-IS). The servers are in a physically secure location on campus and are backed up nightly, with the backups stored in accordance with the AHC-IS retention schedule of daily, weekly, and

monthly tapes retained for 1 month, 3 months, and 6 months, respectively. Weekly backup tapes are stored offsite. The AHC-IS servers provide a stable, secure, well-maintained, and high-capacity data storage environment, and both REDCap and MySQL are widely-used, powerful, reliable, well-supported systems. Access to the study's data in REDCap will be restricted to the members of the study team by username and password.

Each participant will be assigned a unique study identification number by the corresponding research center where the participant is recruited, which will be used for all data contained in the database.

Data entry for responses from participants comprising the general population sample will be conducted at the SDCC after receiving completed forms mailed back. Data entry for responses from participants comprising the clinical population sample and the pregnancy related population sample will be conducted by a combination of SDCC (completed validation survey mailed back, bladder diaries shared over a secure web connection with authentication and data logging) and research center personnel (data from in-person visit, keyed directly into REDCap). SNAP scanning software will be used for data entry from scanned validation surveys and diaries before being imported into REDCap. SNAP is stored on a secure AHC-IS network and access to SNAP will be restricted to the members of the study team by username and password.

VI. Protection of Human Subjects

VI.A. Potential Risks

There are minimal risks associated with completing surveys or participating in interviews regarding bladder health. People may become bored or fatigued during the survey completion and interview process. They may also feel uncomfortable answering some sensitive survey questions. There may be embarrassment or discomfort related to performing the paper towel test or voiding study, however efforts will be focused on creating a private and safe environment to perform the evaluation. Participants will be informed they may discontinue participation at any time.

In addition, there is a risk of loss of confidentiality as well as the possibility that personal information inadvertently may be revealed. Participants will be assigned a unique study number in order to de-identify information collected for analysis datasets. All written surveys and case report forms will be coded with a unique anonymous identifier for tracking purposes and data entered into REDCap using this unique identifier. Electronic surveys will be automatically stored/tracked in the same REDCap database. Data will be stored on password protected servers at the SDCC and all written materials will be stored in locked spaces only accessible by approved research staff. Similarly, all computer files will be secured in password-protected files only accessed by approved research staff.

VI.B. Potential Benefits

It is possible that some or all participants will not benefit at all from the study. Participants may benefit from the knowledge that through research participation they are making a contribution to science. In addition, the participants may learn about their own bladder health and access opportunities for formal consultation with LUTS experts. Women who request additional resources regarding LUTS and bladder health will be provided a resource guide for reference and given contact information for their local providers. All participants will be compensated for their participation.

VI.C. Importance of the knowledge to be gained

The planned study will provide valuable insights into items to incorporate into a bladder health questionnaire. These data are vital to the PLUS consortium objectives and long term goals, including the development of evidence-based interventions to promote women's bladder health and reduce the impact of LUTS. While the individuals in this study will not likely experience direct benefit from participation, the knowledge gained from this work may help other girls, adolescents, and women in the future to prevent LUTS and improve bladder health.

VI.D. Data and Safety Monitoring

Safety monitoring is not applicable given the minimal risk involved in this study. Data integrity and completeness will be monitored as outlined above.

VII. References

1. World Health Organization. *Constitution of The World Health Organization*.; 2006.
2. Lukacz ES, Bavendam TG, Berry A, et al. A Novel Research Definition of Bladder Health in Women and Girls: Implications for Research and Public Health Promotion. *J Womens Health (Larchmt)*. 2018;27(8):974-981. doi:10.1089/jwh.2017.6786.
3. Brady SS, Bavendam TG, Berry A, et al. The Prevention of Lower Urinary Tract Symptoms (PLUS) in girls and women: Developing a conceptual framework for a prevention research agenda. *Neurourol Urodyn*. August 2018. doi:10.1002/nau.23787.
4. Bandilla W, Bosnjak M, Altdorfer P. Survey Administration Effects?: A Comparison of Web-Based and Traditional Written Self-Administered Surveys Using the ISSP Environment Module. *Soc Sci Comput Rev*. 2003;21(2):235-243. doi:10.1177/0894439303021002009.
5. Dillman DA, Sangster RL, Tarnai J, Rockwood TH. Understanding differences in people's answers to telephone and mail surveys. *New Dir Eval*. 1996;1996(70):45-61. doi:10.1002/ev.1034.
6. Vernon SW, Tiro JA, Vojvodic RW, et al. Reliability and validity of a questionnaire to measure colorectal cancer screening behaviors: does mode of survey administration matter? *Cancer Epidemiol Biomarkers Prev*. 2008;17(4):758-767. doi:10.1158/1055-9965.EPI-07-2855.
7. Schwarz N, Strack F, Hippler H, Bishop G. The impact of administration mode on response effects in survey measurement. Jobe JB, Loftus EF, eds. *Appl Cogn Psychol*. 1991;5(3):193-212. doi:10.1002/acp.2350050304.
8. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334. doi:10.1007/BF02310555.
9. Swartz R, de Moor C, Cook K, et al. Mode effects in the center for epidemiologic studies depression (CES-D) scale: personal digital assistant vs. paper and pencil administration. *Qual Life Res*. 2007;16(5):803-813. doi:10.1007/s11136-006-9158-0.
10. Messick S. The interplay of evidence and consequences in the validation of performance assessments. *ETS Res Rep Ser*. 1992;1992(1):i-42. doi:10.1002/j.2333-8504.1992.tb01470.x.
11. Kane MT. Validating the Interpretations and Uses of Test Scores. *J Educ Meas*. 2013;50(1):1-73. doi:10.1111/jedm.12000.
12. Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. *Qual Life Res*. 2018;27(7):1695-1710. doi:10.1007/s11136-018-1815-6.
13. American Educational Research Association. *Standards for Educational and Psychological*

Testing. (Association AP, Education NC on M in, (U.S.) JC on S for E and PT, eds.). Washington, DC: Washington, DC : American Educational Research Association; 2014.

14. Wainer H, Braun HI, Educational Testing Service. Test validity. 1988:267.

15. Streiner D. *Health Measurement Scales : A Practical Guide to Their Development and Use*. Fifth edit. (Norman GR, Cairney J, ebrary I, eds.). New York, New York: New York, New York : Oxford University Press; 2015.

16. Messick S. VALIDITY. *Educ Test Serv Res Rep Ser*. 1987;1987(2):i-208. doi:10.1002/j.2330-8516.1987.tb00244.x.

17. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Helson H (editor), ed. *Psychol Bull*. 1959;56(2):81-105. doi:10.1037/h0046016.

18. Martin JA, Hamilton BE, Ph D, et al. *National Vital Statistics Reports Births : Final Data for 2016*. Vol 67.; 2018.

19. Chin HB, Baird DD, McConnaughey DR, Weinberg CR, Wilcox AJ, Jukic AM. Long-term Recall of Pregnancy-related Events. *Epidemiology*. 2017;28(4):575-579. doi:10.1097/EDE.0000000000000660.

20. McHorney CA, Ware JEJ, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247-263.

21. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. *Br J Obstet Gynaecol*. 1997;104(12):1374-1379.

22. Naughton MJ, Donovan J, Badia X, et al. Symptom Severity and QOL Scales for Urinary Incontinence. 2004:114-123. doi:10.1053/j.gastro.2003.10.059.

23. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *BJU Int*. 1996;77(6):805-812. doi:10.1046/j.1464-410X.1996.00186.x.

24. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). *Am J Obstet Gynecol*. 2005;193(1):103-113. doi:10.1016/j.ajog.2004.12.025.

25. Miller JM, Ashton-Miller JA, Delancey JOL. Quantification of cough-related urine loss using the paper towel test. *Obstet Gynecol*. 1998;91(5):705-709.

26. Gorsuch RL. Common Factor Analysis versus Component Analysis: Some Well and Little Known Facts. *Multivariate Behav Res*. 1990;25(1):33-39. doi:10.1207/s15327906mbr2501_3.

27. Nunnally JC. *Psychometric Theory*. 3rd ed.. (Bernstein IH, ed.). New York: New York : McGraw-Hill; 1994.

28. Armstrong JS. Derivation of Theory by Means of Factor Analysis or Tom Swift and His Electric Factor Analysis Machine. *Am Stat*. 1967;21(5):17-21. doi:10.1080/00031305.1967.10479849.

29. Pett MA. *Making Sense of Factor Analysis : The Use of Factor Analysis for Instrument Development in Health Care Research*. (Lackey NR (Nancy R, Sullivan JJ, eds.). Thousand Oaks, Calif.: Thousand Oaks, Calif. : Sage Pub.; 2003.

30. Gorsuch RL. *Factor Analysis*. 2nd ed.. Hillsdale, N.J.: Hillsdale, N.J. : L. Erlbaum Associates; 1983.