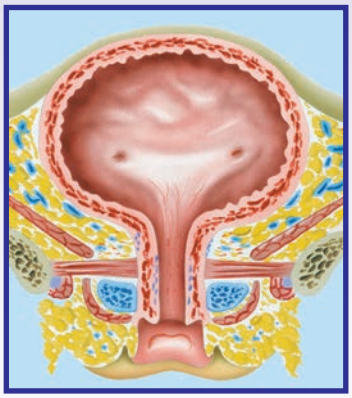
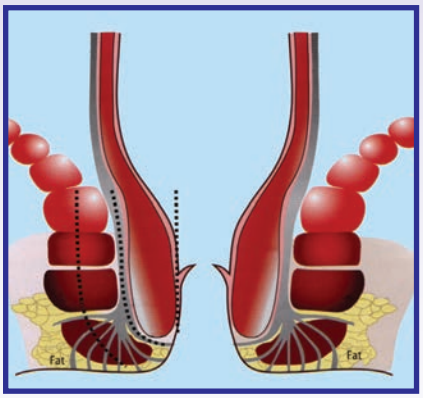
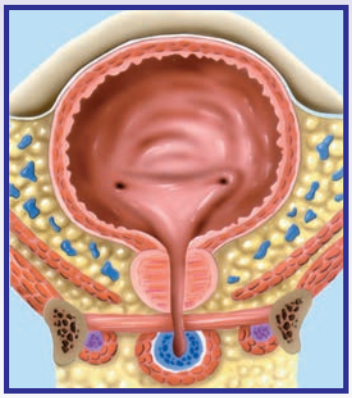
INCONTINENCE

EDITORS

PAUL ABRAMS - LINDA CARDOZO -
SAAD KHOURY - ALAN WEIN







5th International Consultation on Incontinence, Paris February, 2012

European Association of Urology

5th EDITION 2013

ICUD

Committee 23

Research Methodology

Chair

L. Brubaker

Members

I. Nygaard (US)

K. Bo (Norway)

D.G. Tincello (UK)

Y. Homma (Japan)

J. Cook (UK)

M.S. Choo (Korea)

J. Kusek (US)

S. Meikle (US)

C. Payne (US)

CONTENTS

Research Methodology

L. Brubaker

I. Nygaard, K. Bo, D.G. Tincello, Y. Homma, J. Cook, M.S. Choo, J. Kusek , S. Meikle, C. Payne

I. INTRODUCTION

This chapter provides general recommendations for good research practice, including principles of clinical trial design and statistical methodology. In addition, we present specific recommendations applicable to trials for certain types of treatments and certain subgroups of patients.

1. LEVELS OF EVIDENCE

Most contributions to this consultation used the Oxford Centre for Evidence-based Medicine Levels of Evidence and Grades of Recommendation (http://www.cebm.net/levels_of_evidence.asp). However, as these are not an appropriate metric for recommendations in this chapter, the Research Committee used the following categories.

- High: Supported by strong evidence (multiple strong publications)
- Medium: Supported by moderate evidence (limited/moderate level publications)
- Low: Expert/Panel opinion

The report endorses consistent use of the methodology and approved terminology that will not only facilitate incontinence research by producing high quality studies but also facilitate communication about research. Recognized published guidelines produced by the International Continence Society (ICS) [1-15] and Society for Female Urology and Urodynamics (SUFU) [16-18] are useful examples of standardized terminology that advances communication in research.

Research is defined as "systematic investigation designed to develop or contribute to generalizable knowledge" [19] and participants accept risks to advance scientific knowledge and to benefit others [20]. The aim of clinical research in functional urology and urogynaecology is to evaluate treatments intended to prevent and/or significantly reduce symptoms of lower urinary tract dysfunction. The research methodology presented herein is intended to facilitate production of high-impact, high-quality research which will provide evidence needed to inform clinical practice, stimulate other investigators, generate new research ideas, and lead to a better understanding of physiology and pathophysiology of the human health and disease.

Although it is beyond the scope of this section to review the rules of conducting human subject research in depth and extensive coverage is available elsewhere [21-25], all investigators should understand the difference between "research" and "clinical practice". At a minimum, criteria for an ethical research study should meet the following criteria:

- the study should well planned, scientifically sound with clearly defined aims;
- clinical research should be prospectively registered whenever possible;
- the study should be feasible with a realistic chance for completion;
- there should be a reasonable assumption that new knowledge will be provided at the end of the study; and
- there should be an expectation that the results will be published to advance scientific knowledge [21, 26].

Human subject research protocols must be approved by an Institutional Review Boards (IRB), although IRB approval should be regarded only as a minimal ethical standard for research. Ultimately it is responsibility of the investigator to ensure the research is ethically acceptable.

2. PRIMARY GUIDING ETHICAL PRINCIPLES

There are 3 primary guiding ethical principles for human research outlined by the 1979 Belmont Report on "Ethical Principles and Guidelines for the Protection of Human Subjects of Research" [27]. These principles are essential for ethical clinical research and are briefly reviewed in the following section.

a) Respect for Persons recognizes the voluntary nature of research participation and includes informed consent without undue influence.

b) Beneficence requires investigators to maximize

benefits and reduce risks to the subject. The primary concern of the investigator should be the safety of the research participant and careful consideration of the risk/benefit ratio; this includes an ongoing responsibility to monitor research and medical literature as the research proceeds. The investigator needs to critically consider within the expert medical community if there is clinical equipoise for the proposed interventions in their trial. Is one treatment no better than another? Are the research risks reasonable in relation to anticipated benefits? It is the responsibility of the investigator to ensure risks are minimized and potential benefits enhanced as well as that the knowledge gained outweighs the risaks [26]. Of note, invalid research cannot be ethical no matter how favorable the risk–benefit ratio for study participants.

c) Justice requires that the benefits and risks be distributed fairly (i.e. not only using people without access to health care, prisoners, or those impaired). Justice is of particular concern in Phase I testing of pharmacuetical agents and in early investigation of surgical devices/implants. Payment offered for participation in such drug trials may be extremely attractive to poor and disenfranchised subjects. Early device studies may target countries with lax regulatory environment even if there is little intent to market the device there in the long term.

3. INFORMED CONSENT IS AN ETHICAL CORNERSTONE OF HUMAN SUBJECT RESEARCH.

Peer review of protocols by a multidisciplinary team may include members of the scientific community, clinicians, pharmacists, the public/patient groups, the legal profession and individuals who can provide an ethical perspective. Each member of this team reviews the protocol from their particular type of expertise and in doing so aids in safeguarding patient health and well-being.

4. SPECIFIC INFORMED CONSENT IS REQUIRED FOR RESEARCH PARTICIPATION.

The length and depth of detail in consent forms vary widely between institutions. In the extreme, they involve exhaustive pages of information, which explain every alternative treatment with its pros and cons in detail. A general list of requirements for a consent form includes: name of the investigators and contact numbers, a detailed description of the new treatment and its known side effects, rationale for why the new therapy may be preferred to standard therapy. A summary table of the results of previous studies using the drug can be helpful when available. A statement that the patient may decline to be in the study with no subsequent consequence to their ongoing medical care is generally provided and whether or not remuneration is expected. Additionally, there should be a statement about payment for medical care required during the course of the study if there is an adverse event associated with the intervention. An understanding that the patient will be randomly assigned to treatment should be included when relevant, written using terms that are meaningful to potential participants [28-29].

5. THE INVESTIGATOR HAS AN ETHICAL RESPONSIBILITY TO TAKE RESPONSIBILITY FOR ALL ASPECTS OF THE RESEARCH.

This will insure that the work is done rigorously and to maintain the integrity of the research [30]. International Committee of Medical Journal Editors (ICMJE) now requires that information about trial design be placed into an accepted clinical trials registry prior to participant enrollment [31]. For more information, see more extensive information available elsewhere [30, 32-35]. A list of registries acceptable to the ICMJE can be found on their website. One objective of trial registration is to reduce publication bias given that trials with positive findings are more likely to be published than those with negative findigs. An important feature of one registry, clinicaltrials.gov, is a results database [36].

Consider this pragmatic, simple test to use when resolving an ethical research dilemma: “Imagine what you are preparing to do will be reported the next day on the front page of your local newspaper. If you are comfortable having colleagues, friends, and family know what you did, chances are you acted responsibly”.

6. ENSURING PARTICIPANT SAFETY IS PARAMOUNT.

An independent data safety monitor (DSM) or data safety monitoring board (DSMB) is important to evaluate the study on an ongoing basis to determine early evidence of significant harm or benefit [37-38]. Depending on the size, complexity, and risks of a trial, the DSMB is comprised of experts needed to monitor interim data to ensure the safety of the participants. The DSMB should be established prior to initiation of the trial. In addition to reviewing results of the study for safety monitoring they may evaluate interim analyses to ensure that a treatment is not producing unacceptable levels of side effects and/or efficacy [29, 37]. A priori stopping rules or boundaries are established to assess if the study should continue or be terminated due to futility (that is, no conclusion will be drawn due to low enrollment, few outcome events, or high drop out rates etc.), reaching an endpoint, or identifying increased risks. Guidelines for stopping the study should be agreed upon, prior to the start of the trial [30]. Interim analyses (in particular those based on efficacy) will have implications for the study power. Specialist statistical advice and support will be essential to address these issues [39, 40]. Investigators must not be aware of the results of interim analyses, however, since this may cause bias by influencing how vigorously any given patient is recruited into or followed up in the study, and most importantly, runs the risk of a type

II error (ie mistakenly concluding benefit when there is none). Nevertheless, emergency procedures for unblinding a patient's allocation are required in case of a severe side effect or concomitant serious illness where knowledge of treatment assignment is essential for patient management and safety.

7. HIGH QUALITY DATA MANAGEMENT IS KEY TO PROVIDING VALID AND ETHICAL RESEARCH RESULTS [41].

In addition, ethical conduct of research includes timely and complete reporting. Clear guidelines for authorship have been established by medical journals [42]. Beginning with the research contract, authorship rules should be established according to accepted guidelines. In general, authorship credit should be based on 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published. Authors should meet conditions 1, 2, and 3.

8. USEFUL WEBSITES:

- The National Reference Center for Bioethics Literature (http://bioethics.georgetown.edu/nrc/index.htm)
- International Bioethics Organizations Database (http://bioethics.georgetown.edu/databases/Organizations/index.htm)
- International Committee of Medical Journal Editors (http://www.icmje.org)

9. FINANCIAL CONFLICTS OF INTEREST

Many investigators are involved in testing new drugs or devices developed by industry and have the potential to significantly supplement to personal income. It is acceptable for investigators to receive contracted financial support to perform this research and a principled partnership between industry and investigators is essential if we are to preserve medical progress [43]. Financial conflict of interest policies have been developed due to ethical concerns about potential biases that may influence trial design, conduct, over interpretation of positive results or not publishing negative results [30, 44]. It is important that investigators do not receive money directly (personal income) from industry sponsors but rather through a research contract through an appropriate entity with research oversight capabilities. Disclosure of these relationships helps maintain scientific integrity and preserve public trust in the scientific process. This is a rapidly evolving area and investigators are encouraged to clearly understand and disclose such relationships.

An investigator's institution may have additional definitions and reporting requirements for financial conflict of interest disclosures. There are many potential relationships between physicians and industry; it is preferable that the nature of the relationship and its financial magnitude if any, be fully defined rather than categorized (i.e. "consultant") so that the reader can appropriately assess the disclosure.

II. DEFINING THE RESEARCH QUESTIONS

The investigator(s) should take a deliberate approach to formulate the specific research question, based on a careful review of related clinical research and relevant studies that are well designed and clinically relevant. The research reviews provided by the Cochrane Incontinence Group (http://healthsci.otago.ac.nz/dsm/wch/obstetrics/cure) provide an excellent starting point for most major incontinence topics. Based on a thorough literature review, the investigator clearly describes the primary research question(s), summarizes the background information, and formulates the rationale, objectives and hypotheses for the study. The investigator(s) should formulate the simplest study design which will provide the highest quality of evidence to test the given hypothesis in a cost and time-efficient manner. Whenever possible, basic or translational research should occur as part of the clinical research study in order to advance discovering of underlying mechanisms or pathophysiology. This balance between breadth and depth greatly increases the yield of the research effort [45].

III. EXPERIMENTAL STUDIES

1. RANDOMIZED CONTROLLED TRIALS

Experimental studies have the potential to provide a higher level of evidence than observational studies. The randomized double-blind clinical trial (RCT) is considered the gold standard study design. Properly planned and executed, the RCT is the optimal approach to limiting allocation bias [46]. The study participants are assigned to a treatment group by a random (chance) mechanism that ensures adequate concealment so that neither the study participant nor the investigator can influence treatment group assignment in order to reduce allocation bias. Subject assignment must be concealed during enrollment (for example, by separating allocation from the process of recruiting subjects, and by using remote randomization such as by telephone or web-based procedures), and wherever possible treatment allocation must be concealed during the trial (for example, using blinding with or without placebo). In order to minimize bias, the randomization process must be concealed from those recruiting subjects to the trial [48- 49). This can be achieved most effectively by the use of central telephone randomization.

In drug studies, a pharmacy can maintain identical treatment drug and placebo already randomly allocated into individual subject portions. These are distributed consecutively as subjects are enrolled in the study. In some studies, blinding of subjects and health care providers may not be possible, for example in trials of some surgical procedures or health care delivery methods. In almost all cases, however, the personnel collecting outcome data should be unaware (blinded) to the subjects' treatment allocation. RCTs require ethical equipoise and are usually expensive to conduct optimally.

The purpose of randomization is to produce groups that are, on average, comparable. A per-protocol analysis retains this property only in the unlikely situation when non-compliance is unrelated both to the patient's underlying state of health and the treatment received [50]. The intention-to-treat approach in pragmatic trials retains the full benefits of randomization and has the advantage that the comparison will more closely reflect the relative effectiveness of the treatments when applied in real clinical practice, where non-compliance is a common occurrence [51].

a) Simple randomization can use computer-generated random numbers, either prepared specifically for the trial or using existing tables of random numbers where the digits of 0-9 appear with equal likelihood in each entry. Treatments are assigned to odd or even numbers. As the total number of subjects in the trial increases, the balance of numbers and characteristics of subjects between the groups improves. In small trials, however, balance is not assured by simple randomization. Appreciable imbalances in subjects per group may be particularly important in a multicenter study where imbalances in assignment can occur within individual institutions.

b) Block randomization is one method used to prevent imbalances in subject numbers assigned to each group, particularly when the number of subjects in the trial is small. With block randomization, the total sample size is divided into blocks of a given size. Within each block, the group is assigned so that there are equal numbers allocated to each group. To prevent investigators from learning the block size and being able to guess order of assignment, the block size can be varied, usually at random from a small number of alternatives. In any case, blocking prevents serious imbalances in characteristics across groups when used in conjunction with stratification as described below.

c) Stratified randomization

Most disease states have factors known to influence the outcome of treatment, for example symptom severity or gender. A form of randomization that accounts for such factors is called stratified randomization [46-47]. Stratified randomization ensures equal distribution of subjects with a particular characteristic in each group when blocking is employed within strata. Stratification is usually restricted to a small number of factors, in particular those most likely to influence outcome. Despite its complexity, stratified randomization is usually helpful in a multicenter trial, so that both the numbers of subjects in each group and the important factors influencing the outcome can be balanced within each site. An alternative method exists to cater for more factors at once, known as minimization, where the characteristics of individuals already randomized alter in a systematic manner the chances of a given subject being allocated to the different trial groups, so as to maximize the resulting balance of these factors [46-47, 52].

Although the classical RCT involves the study of parallel groups, other options are possible and may overcome some of the limitations of the classical approach [53]:

d) Parallel Group Trials

This design generally includes one group of subjects assigned active treatment and a second (parallel) group a placebo or another treatment not believed to be an active treatment. The key feature of this design is that both groups (treatment and comparison group) are assembled and followed at the same time. In clinical trials of drugs the dose of the drug tested may be either a single dose or multiple doses to determine clinical benefit and/or minimize side-effects. More complex study designs can in some circumstances be worth considering – for example, factorial trials where two or more interventions can be investigated simultaneously [54-55], and cluster randomized trials whereby groups of participants (defined by some common feature, e.g. members of the same health maintenance organization, clinic, etc.) rather than individuals are randomly allocated to the trial arms [56]. This strategy might be employed when studying an intervention requiring policy changes in an institution, with a hospital, clinic, or health care system being the unit of randomization.

e) Crossover Trials

Subjects receive both the treatment being studied and the placebo/alternative treatment, with random order of treatment assignment. The benefit of crossover studies is that they eliminate the effect of variation between groups of participants seen in parallel trials, that is each subject serves as his/her own control. Crossover studies are particularly well suited for small studies, where the course of the disease under study is believed to be stable, and where the primary objective is to measure a short-term change in the outcome (e.g. urinary symptoms) in response to treatment. The duration of treatment required before the anticipated effect may be observed is critical in determining whether the crossover study design is appropriate – too long a period time before an effect becomes evident, and the disease state may vary before the study

participant has completed all arms of the trial; too short, and it may not be possible to detect the effect during the period of data collection. Carryover of treatment effects from one treatment period (or observation period if a placebo follows treatment) to another, make this a challenging study design to implement in many cases. To avoid a carryover effect, a washout period should be included, in which participants receive either placebo or no treatment. A run-in period in which signs and symptoms of the illnesses studied are monitored may be necessary before treatment begins to ensure that only those whose disease state is stable are entered into the study. Given all thesefeatures and limitations, this design is unlikely to be widely applicable in studies of interventions for incontinence.

There are two types of clinical trial designs that show the similarity of medical treatments: non-inferiority trials and equivalence trials.

f) Non-inferiority trials

the primary objective of a non-inferiority trial is to demonstrate that the new treatment is not unacceptably worse than that of the standard treatment. This design may be appropriate when one treatment is less costly , may offer advantages related to improved quality of life, or has fewer side-effects, and the current belief is that both treatments have a a similar effect on the disease or condition of interest . A key feature of this type of trial design is the ability to rule out a non-inferiority margin; that is, a minimum threshold for an unacceptable loss of efficacy [57-59]. As an example, if prior independent studies found similar effects of pelvic floor exercises and a drug in the treatment of stress incontinence, one might set up a study to demonstrate that no important difference in clinical effect was present between the two treatments by direct comparison. Currently there is considerable uncertainty about the design of non-inferiority trials as it relates to U.S. and European drug regulatory bodies [60-62]such that there have been recommendations to consider alternative designs [57]. Non-inferiority trials have been infrequently used in studies of urinary incontinence. A recent example is the Value of Urodynamic Evaluation (ValUE) trial, a non-inferiority randomized trial of preoperative urodynamic investigation in women undergoing stress urinary incontinence surgery [63].

g) Equivalence trials

The goal of an equivalence trial is to show that the experimental treatment is not more inferior and not more superior to a standard treatment by a given amount [64-66]. The clinically important difference between the standard treatment group and the new therapy is established prior to start of the study (Fleming TR 2000), and it is shown, by formal statistical testing, that the difference between the control and study treatments are not large in either direction (more beneficial or less beneficial). A recent trial of midurethral sling surgery illustrates this type of study design [67].

h) Superiority trials

The most common study design for clinical trials of urinary incontinence and for many other diseases/ conditions is the superiority trial where the goal is, in its most simple form, to show one treatment is better than another (or no treatment, placebo). The general approach is to establish a priori a clinically meaningful effect that you wish to detect. Once conducted, determining superiority requires that the findings permit the investigator to reject the null hypothesis that the two distributions of the treatment effect are equal, in favor of the new treatment being better than the control [58].

2. NON-RANDOMIZED CONTROLLED CLINICAL TRIAL(S)

This category includes trials in which treatment allocation is known to the investigator prior to obtaining informed research consent (for example, day of clinic appointment). This approach often results in baseline characteristics of the treatment groups being significantly different, decreasing the ability to interpretate trial results. Because of this major shortcoming, this type of study should be used rarely, if at all.

3. PRAGMATIC AND EXPLANATORY TRIALS

There is an important distinction between pragmatic and explanatory trials [68- 69], and correspondingly, between intention-to-treat and per-protocol approaches to data analysis [50, 70]. This distinction has a number of facets. In pragmatic trials the interventions are designed to be as close as possible to treatment options in clinical practice (including multiple patient management choices) and entry criteria are usually relatively liberal in comparison with explanatory trials. In addition, pragmatic trials may involve a wide variety of outcome domains, including patient-completed questionnaires, and an economic evaluation of outcomes. As a result of intention-to-treat data analysis, pragmatic trials will tend to yield lower estimates of treatment differences than explanatory trials. It may be of interest to gauge the effect of treatment given full compliance; therefore, full data analysis ideally incorporates both intention-to-treat and per-protocol approaches [50]. The primary analysis, though, should follow the intention-to-treat principle.

Data from pragmatic trials are analyzed by intention-to-treat, according to the group to which subjects were randomized, regardless of the extent of compliance with the intended treatment. In explanatory trials, data are analyzed accounting for compliance. This per-protocol approach may exclude serious non-compliers, analyze data according to treatment actually received, or allow for degree of compliance in a statistical model.

At first sight, the explanatory approach appears more attractive. However, there are considerable limitations to the explanatory approach, particularly when the intention is to draw inferences from the trial to wider clinical practice (generalizability).

4. DRUG TRIALS ARE CATEGORIZED ACCORDING TO THE FOLLOWING DEFINITIONS [30, 50].

a) Phase I studies

The first studies of a drug in humans, often open label and uncontrolled, concentrating on safety and frequently but not exclusively carried out in healthy volunteers. Pharmacokinetic and tolerance information is obtained from Phase I trials.

b) Phase II studies

The first attempts to investigate treatment efficacy, often the first use of the drug in subjects and focusing on short-term outcomes. A common objective of Phase II studies is dose finding in terms of efficacy. Two sub-types may usefully be distinguished: Phase IIA studies where single treatments are considered in relation to a minimum response prior to further investigation; Phase IIB where direct comparisons are made between interventions, albeit on a small scale and not necessarily involving randomization [71].

c) Phase III studies

Large-scale, authoritative randomized studies performed once the most likely effective and tolerated treatment regimens have been established. The objective is often to establish that the intervention is suitable for registration/approval with the appropriate regulatory authority. Trials are conducted after submission of a new drug application (NDA), but before the product's approval for market launch. Phase IIIB trials (between submission for approval and receipt of marketing authorization) may supplement or complete earlier trials, or seek different kinds of information (for example, quality of life or marketing). Phase III trials are also used to investigate the effectiveness and cost-effectiveness of various interventions – that is, non-drug including organizational issues – and not necessarily with reference to regulatory authorities. All Phase III trials should be subject to a formal sample size calculation – for instance to obtain sufficiently precise estimates of the comparisons between treatments or to have a reasonable chance (power) of detecting a difference if one exists (see section II C 7 below).

d) Phase IV studies

These investigations are usually carried out after registration/approval, to investigate the drug's safety and efficacy in different populations. Such post-marketing surveillance studies are typically larger and simpler than regulatory studies; they may lack a control group and are often conducted using surveys.

5. BIAS, BLINDING AND EFFECTS ON VALIDITY

Bias can be introduced at many stages of a study including patient selection, randomization, assessment of outcomes, and statistical analysis and interpretation. Bias occurs because of previously conceived ideas held by those involved, which consciously or unconsciously affect their actions and observations. In addition to observer bias, an amount of observer error is inherent in outcome measures that require clinical interpretation.

a) Blinding

In order to minimize bias, research design should strive for the highest practical level of blinding. Blinding is the process by which key elements of knowledge are withheld that can otherwise lead to bias. Blinding should not be confused with concealment of allocation, referring to withholding knowledge of assignment in advance, which is a prerequisite for the validity of any trial [48, 72]. While blinding is important, its effect is lower than that of concealment of allocation [73]. In most drug studies placebo pills allow blinding of all relevant participants.

b) Unblinded trials

They are conducted in an open manner where both subjects and study investigators are aware of which treatment has been assigned. While certain types of therapy may require investigation in this manner (e.g., some surgical trials), there remains considerable opportunity for bias. Both subjects and investigators may have preconceived ideas regarding the benefits of a particular treatment that can influence the reporting of symptoms and/or their outcome.

c) Single blind trial

Commonly, in a single blind trial, 'single blinding' may refer to blinding the outcome assessor (but not the subject); this is common in trials in which blinding the subject is not possible (such as pelvic muscle exercise trials). It is important to specify who is blinded to which aspects of study execution, rather than using only the vague term "blinded".Hence, in these studies the only possible blinded person is the assessor. It may be advantageous for the clinical staff to be aware of the assignment to allow them to monitor the health and safety of individuals, since the potential effects of the treatment (side effects) will often be known in advance. Single blinding ameliorates biased reporting of symptoms and/or side effects by subjects. However, clinical staff can influence data collection and change other aspects of subjects' care when they know which study treatment subjects are receiving.

d) Double blind trials

In double blind trials, both parties who could influence outcome are unaware of group assignment. Often this is just the subjects and the clinical team responsible for their care. Selection of an appropriate external comparison group is challenging and factors unknown or not measured by the investigators may adversely influence the findings.

IV. ELIGIBILITY CRITERIA

Eligibility criteria impact every aspect of a study's progress, from recruitment to analysis to acceptance of the study's results by the community. Yet, in the planning phase of a study, this vital aspect is often given short shrift. Eligibility criteria will differ, depending on the main goal of a study. For a Phase I trial, designed to understand safety of an intervention, eligibility criteria should be very narrow, and should focus on minimizing risk and maximizing ability to demonstrate proof of concept. A Phase II trial, designed to test the efficacy of a treatment in a fairly ideal population, will have broader eligibility criteria, but will exclude individuals in whom it is highly unlikely a treatment will be helpful, or highly likely that a treatment might be harmful. If a treatment shows promise at this point, eligibility criteria for a Phase III trial are as broad as possible, to test the effectiveness of the treatment in as real world of a setting as possible, thus maximizing generalizability. Conclusions can only be as broad as the eligibility criteria allow. Thus, a drug found to be effective in a study that included only healthy middle-aged people can be recommended for that group, but not for older people, children, people with medical conditions, or other groups excluded by the eligibility criteria.

While some **exclusion criteria** are necessary to reduce risk to potential participants, it is important to include individuals in whom the treatment would likely be used once widespread. For example, in testing a new drug for urgency incontinence, that is excreted by the kidneys, it is important to exclude people with high serum creatinine levels, but equally important to include older individuals with multiple co-morbidities in specific studies, as these are also likely to benefit from a drug for urgency incontinence once released.

Particularly in longer-term studies of pelvic floor disorders, some exclusion criteria are needed to maximize potential for follow-up. Such studies often exclude people likely to move or die during the study period, people with conditions that impair adherence (such as drug or alcohol dependence, severe mental illness or homelessness) or factors that make follow-up more difficult (such as lack of a telephone). For ethical reasons, it is important to either exclude people with significant cognitive impairment (such that informed consent is impossible) or to plan a consent process with a legally authorized representative in mind. However, all exclusions impact the generalizability of the conclusions and so such trade-offs must be carefully thought through.

Additionally, inclusion criteria related to the condition of study must be considered in concert with the consideration of the primary outcome measure. If an inclusion measure defines "disease", it is essential that the outcome measure does not include this same measure or cut-point as "success". For example, if a surgical trial comparing two treatments for pelvic organ prolapse defines prolapse as >/= stage II (and excludes from enrollment only women with </= Stage I prolapse) and the primary outcome measure defines successful treatment as prolapse at or above the hymen, then women enrolled with maximal prolapse at points 0 or -1 may be considered successfully treated when in fact, their condition has subjectively not changed. It is also important to note that a given condition, such as stress urinary incontinence, will be present or absent depending on which outcome measure is used to define it. Avoid using one outcome measure to exclude women from a study and then a different outcome measure defining the same condition as the primary outcome measure.

V. SAMPLING STRATEGIES

Every study must address its sampling strategy, that is, the selection of participants. In principle, sampling should involve random selection. In practice, however, this ideal is rarely met outside of large-scale epidemiological studies. For example, RCTs are drawn from a subset of the population, often limited to those with access to academic medical centers plus the willingness and ability to participate. Where this is the case, it is crucial to provide descriptive information about the study sample, so that its representativeness can be judged. Guidelines for reporting of RCTs include requirements to state the study population, give details of inclusion and exclusion criteria, and present clearly the numbers of eligible subjects who were not randomized and the reasons [32, 48- 49].

A study may require a sample that is representative of the community overall or one representative of patient groups suffering the condition/disease. In principle, this is achieved by taking a simple random sample from a known population. In practice, a list of all eligible individuals is obtained and then a sample is drawn by a method in which each member of the population has an equal probability of selection ('epsem'). Even in ideal circumstances, however, some sophistication on this basic method is usually desirable or necessary. For example, in stratified sampling, subjects are arranged into subgroups and the sampling is performed within each subgroup separately. This ensures that the sample is representative of the population in terms of these subgroup character-

istics. In multi-stage random sampling, the population is first divided into primary sampling units (such as hospital, health center, or surgeon), and a sample of primary units is selected. The secondary sampling units, usually individual subjects, are selected within the primary sampling units. A special case of multi-stage random sampling is cluster sampling where all individuals within each primary unit are included. Standard procedures for sampling should be followed [56, 74].

It is important to note that, while the technicalities of random selection of subjects for a study are closely related to the random allocation of subjects in an RCT (and indeed there are similar issues in trials relating to stratification and clustering)[56], there is an important distinction in the objectives of the two procedures. First, the (ideally random) selection from the population of eligible subjects concerns the external validity or generalizability of the study findings (RCT or otherwise). Independent of this, the random allocation of subjects in an RCT is concerned with the internal validity or comparability of the trial groups. It is usually obvious whether the study can be performed at a single institution, or whether a multicenter study will be required. Single institution studies have the benefit of being less complicated from a logistical perspective. While multicenter trials are more complex to manage and are usually more expensive, they provide larger numbers of participants in a shorter period of time, and increase the generalizability of research findings.

VI. DATA COLLECTION

Although the number of questions in a single study should be limited, it is still relevant to record as many as observations as is possible without jeopardizing recruitment or retention with onerous demands. Strong consideration should be given to collection of the following data groups [16],[75].

a) Baseline data:

b) Observations:

1. Patient's observation/Subjective measures
2. Clinician's observation/Objective measures

c) Tests

1. Quantification of symptoms—bladder diary, and
2. Pad tests for incontinence outcomes
3. Urodynamics

d) Follow-up

e) Quality of life measures

f) Socioeconomics

It is useful to consider collection of the following variables:

- Age
- Race/ethnicity
- Gender
- Body mass index (height and weight)
- For women – obstetric history, including parity and menopausal status
- Smoking status
- Co-morbidities such as cardiovascular, respiratory and neurologic conditions and diabetes
- Medication use
- Past surgical history
- Level of education
- Mental and physical status
- Prior treatment for pelvic floor disorders, including behavioural, pharmacological and/or surgical interventions
- Sexual function

VII. OUTCOME MEASURES

Specific discussions of the most appropriate outcome measures for particular studies of incontinence and pelvic organ prolapse are discussed elsewhere in this consultation. The purpose here is to define the general concepts of primary and secondary outcomes, which are relevant to both sample size determination and data analysis. The distinction between these two sets of outcomes depends on the context of the trial, and should be decided at the planning stage of the study. Primary and secondary outcomes should not be confused with the distinction between primary and secondary analyses of trial data.

1. PRIMARY OUTCOMES

Primary outcomes are those viewed by the researchers to be of central interest. Sample size calculations are based on anticipated changes in the primary outcomes. The final study population may or may not be sufficient to allow robust comparisons of secondary outcomes. Trial results that are based on the primary outcomes can lead to reseponsible alterations in standards of patient care.

The number of primary outcomes in a particular trial will depend on the nature of the interventions and the number of independent domains. The number of primary outcomes is usually limited to three. Sample size calculation is based on the primary outcomes. The number and nature of outcome domains in a particular study will vary depending on the study's perspective. Careful selection of outcome measures

that are valid and clinically relevant is intrinsic to the success of research; usually no single measure can fully express the outcome of an intervention.

2. SECONDARY OUTCOME

Secondary outcomes can be relatively large in number. They are not the focus of the main study objectives and are rarely used directly in sample size estimation. Secondary outcomes are exploratory, i.e., as hypothesis-generating exercises for which independent confirmation is essential or used to monitor: e.g. safety, side-effects, costs or patient acceptance.

3. TYPES OF OUTCOME MEASURES

Traditionally, a patient's success after a particular treatment was reported by the physician, using subjective information that differed from physician to physician. Given the mismatch between physicians' and patients' perceptions of treatment effectiveness, this type of physician reported outcome measure has been largely replaced with patient-reported outcomes (PROs). The Food and Drug Administration defines a PRO as any report of the status of a patient's condition that comes from the patient, without interpretation of the response by the clinician or anyone else.

(http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf)

There is emerging consensus that PROs are most appropriate when describing success or failure of therapies. However, physician-reported outcomes can yield valuable information about mechanistic effects of therapies and are also vital in developing predictive data for future use. As an example, a patient's report of a bulge is a reasonable outcome measure for a study of surgical treatment for POP. However, if, over time, women with stage II POP after treatment who were initially asymptomatic became symptomatic, it would be pertinent to include prolapse stage, as an anatomic outcome measure, in the definition of success.

Other types of physician reported outcome measures (often called "objective outcomes"), used frequently in pelvic floor disorder research, include results obtained from urodynamic testing (e.g., presence or absence of detrusor overactivity incontinence), radiologic imaging (e.g., various angles measured during pelvic ultrasonography), physical examination (e.g., the POP-Q system for assessing pelvic organ prolapse), laboratory values, biomarkers, etc. , as well as tests collected by the patient but interpreted by the physician (such as pad tests and bladder diaries).

The primary types of PROs include [76].

- **Disease/condition specific**: measure patients' perceptions of a specific condition. These cannot be administered to people without the condition under study.
- **Population specific:** specific to a certain population (for example, women or men or elderly) with or without a given condition.
- **Dimension specific:** assess one aspect of health status. The primary dimensions studied include:
  - Physical function (e.g., activities of daily living)
  - Symptoms (e.g., urinary incontinence)
  - Psychological well-being (e.g., depression)
  - Social well-being (e.g., sexual function)
  - Cognitive function (e.g., mental status evaluation)
  - Role activities (e.g., household or work activities)
  - Personal constructs (e.g., body image)
- **Generic measures:** measure broad aspects of health which can be used to compare across populations (The SF- 36 is a commonly used measure). These are often less responsive to clinically significant changes in health than condition-specific instruments
- **Utility measures:** (e.g., EuroQol) incorporate preferences or values attached to individual health states. This type of instrument produces evidence for the overall value of health states to society and can be used in cost-utility analysis.
- **Summary items:** participants summarize an overall picture of their status in one item (for example, the transitional item queried in the SF 36: "Compared to one year ago, how would you rate your health in general now: excellent, very good, good, fair, poor?"
- **International Classification** of Functioning, Disability and Health (ICF), WHO 2001
- **Unified and standard language** and framework for discription of health and health-related states
- **Body, Individual, Society**
  - Body functions and structures
  - Activities and participation
- **Body functions:** physiological and psycological functions of body systems
- **Body structures:** anatomical parts
- **Impairments:** problems in body function or structure such as significant deviation or loss

- **Activity:** execution of a task or action by an individual
- **Participation:** involvement in a life situation

While not yet in wide use, web-based resources for administering computerized adaptive tests are being developed. One example is the Patient Reported Outcomes Measurement Information System (PROMIS) system (www.nihpromis.org). Using this type of testing, researchers can lower participant burden substantially. For example, in a measure of physical function, someone who endorses inability to dress themselves would not be asked the next series of questions in a typical validated physical function questionnaire about walking or exercising; someone who endorses ability to run marathons would not be asked questions about less strenuous activities. Nascent work has begun on the use of the PROMIS system in pelvic floor disorders [77].

Investigators should strive to use validated outcomes measures whenever possible. A fully validated outcome measure is demonstrated to 1) be reproducible when administered to the same person twice, 2) yield results that align with other gold standard measures (content validity), 3) change with successful treatment (be responsive to change), and 4) measure what it is intended to measure (face validity). (A full discussion of psychometric properties needed to validate a questionnaire can be found elsewhere in this consultation.) Much progress has been made in this area over the last decade, and therefore, when planning a study, the default position should always be choosing a validated outcome tool as a primary outcome measure. However, it is important to note that just because a questionnaire is "valid" does not mean it can be used in any population in any language. Researchers must take care to use instruments in the populations for which they were intended.

Adverse event recording and analysis is an important aspect of outcome measures, especially when the benefits may be offset by harms. Adverse event assessment should be systematiclly included in every interventional study. The Clavien-Dindo system for structured reporting and analysis of surgical complications has been used in several studies comparing pelvic floor interventions[78]. In this system, complications are classified into one of four categories based on the type of therapy needed to correct the complication.

The timing of outcome assessment should reflect the study question. Studies reporting efficacy of incontinence treatment should follow subjects for a minimum of one year. Although early follow-up can provide important insights about adverse events, a great deal more emphasis needs to be placed on long term effectiveness. The follow-up time for a trial should be at a fixed time (for logistical reasons, this is in practice often a short time window) relative to randomization rather than when treatment was actually received, since again this is the only way of ensuring a valid comparison. The planned timing of follow-up at a fixed time relative to randomization should, however, allow for any likely delays in receiving treatment, e.g., due to surgical waiting lists.

VIII. SATISTICAL CONCERNS

1. SAMPLE SIZE CONSIDERATIONS

There is no single answer for sample size determination; often the calculation proceeds around a 'circle of specifications' (involving, say, power, targeted difference in effect size and population available for study). Furthermore, the ideal of the target being the minimum for clinical significance cannot always be met; rather, the aim in practice is to produce a convincing argument (among the researchers themselves, and also to funding bodies and regulatory agencies) that the sample size has an adequate chance of detecting differences that are (a) feasible, and (b) worthwhile detecting in clinical terms. A common failing is selecting a target difference that is too large, often derived from differences that have been observed or published previously rather than based on considered clinical judgment. Preliminary investigations into the levels of treatment effects that patients themselves consider worthwhile should be carried out much more commonly than is the case at present.

Sample size should be calculated in the planning stage of all studies. There are many formal equations to assist in this process, details of which will not be given here [70-71, 79-80]. Rather, the emphasis for this discussion is on the concepts involved and the information required for the calculations to proceed.

2. SAMPLE SIZE CALCULATION

One approach to sample size calculation is based on the required precision of an estimate, which is relevant to both descriptive and analytical investigations. The basic issue is one of precision (measured by the standard error, SE) or margin of error (which depends on the SE but is more specifically defined as half the width of the 95% confidence interval [CI] around the estimate). The higher the level of precision specified in advance (i.e., the smaller the SE and the narrower the CI), the larger the sample size will need to be. However, the margin of error depends on the nature of the primary outcome variable, i.e., whether it is a continuous variable (such as maximum urinary flow rate) or a binary variable (such as the presence or absence of self-reported urgency incontinence). For a continuous variable, the variability (standard deviation) of the measure must be estimated for relevant subjects; this may

be derived from some combination of clinical experience, the literature, or a pilot study. The larger the variability, the larger the sample size required. For a binary variable, its prevalence must be estimated in the population to be studied, since the SE for such variables depends on their prevalence.

A second approach, based on power, is the most commonly used and requires that the study have adequate probability (power) of detecting a given (target) magnitude of effect. It requires similar prior information, including estimates of the variability for continuous measures and the magnitude of proportions for binary variables. In addition, it requires specification of three other quantities: the significance level, the power, and the target difference. Significance is a statistical term that tells how certain one can be that a difference or relationship exists. It does not necessarily imply that the result is clinically relevant, just that the result is likely to be accurate. The significance level, termed alpha, is conventionally, though not necessarily, set at 5%. Power is defined as the probability that the study will detect (as statistically significant at the alpha level specified) a given target difference between the groups, if such a difference exists. Power is commonly specified in the range of 80% to 90%, which implies a risk of not detecting the target difference of between 20% and 10%, respectively. For a trial involving anything other than minor risks and expenditure, a power closer to 90% than 80% would seem preferable [32], which leads to a larger sample size (as does a stricter alpha level of, say, 1%). This is most pertinent when a lack of statistical significance is obtained in a small trial, particularly when the sample size was not planned using a power calculation [47]. This is the basis for the adage that "the absence of evidence is not evidence of absence"[25]. A planned unequal allocation to the trial groups also requires an inflation of the sample size [47], as does interim analyses. By multiplying the number of significance tests performed, studies with interim analyses generally require stricter significance levels at each analytical point [50, 70].

3. THE TARGET DIFFERENCE

This is the last, and arguably the most important, quantity that must be specified for the power-based approach to sample size calculation. The target difference is defined as the minimum difference between treatment groups considered to be clinically significant. Clinical significance is an entirely different concept from statistical significance. Investigators must estimate the clinical significance as the magnitude of difference (in means or proportions) that would lead to a change in clinical management for the target group of patients. For example, a study might propose that a 20% difference in incontinence episode frequency is a clinically meaningful response. Ideally, such an assumption would be based on surveys of patient behavior but in practice the decision is often based on clinical judgement. In any case, the smaller the clinically significant target difference, the larger the required sample size. Statistical significance means that the observed difference, whatever its magnitude, cannot reasonably be considered as being due to chance. Statistical significance (denoted by the p-value) represents the strength of evidence against the null hypothesis [81]. The degree of clinical significance can be inferred only with the additional information of a confidence interval for the comparison between groups. A very large trial may achieve a high level of statistical significance with a very small effect size and therefore be of little clinical significance. A trial designed in such a way poses ethical challenges because many extra subjects are exposed to risk without meaningful benefit.

With each approach, appropriate adjustment for attrition (loss to follow-up) should be performed. This is commonly achieved by simply increasing the planned sample size in proportion to the anticipated attrition (i.e. to predict the reduced effective sample size that will be available for the analysis).

IX. ANALYSIS

The analytic plan should be consistent with study aims and the a priori analytic plan should be included at the time of the protocol entry into a clinical trials registration system. Currently, high tier journals will often request the statistical analysis plan and protocol at the time of the review of the report. This section will not contain any technical details of statistical methods, which are available in standard texts [47, 82-83], but rather will summarize concepts of data analysis.

It is established practice that the **primary analyses** (for both primary and secondary outcomes) of an RCT should be on an intention-to-treat basis [48, 72]. **Secondary analyses** incorporating non-compliance and/or which treatment was actually received may be justified in addition to the primary analyses. In practically all situations, hypothesis tests should be two-sided (i.e., allowing for the possibility that the difference could have been in either direction, that is benefit or harm), rather than one-sided. One-sided tests are only appropriate if a difference in one direction is not just unlikely, but would not be of interest, such a demonstrating only superiority of a new drug.

Regardless of the type and complexity of statistical techniques used in analysis, the general underlying principles behind hypothesis testing and estimation apply. In particular, the statistical significance of a hypothesis test should be interpreted critically. The actual p-value should be considered, rather than just whether or not it is below an arbitrary threshold such as 5% [48]; indeed, the p-value is better considered a measure of the strength of evidence against the null hypothesis, on a continuum or 'shades-of-grey

[83-84]. The direction and magnitude of the trial comparison should be presented with an appropriate confidence interval to indicate the possible clinical significance and precision of the comparison [48, 85].

Appreciable **loss to follow-up** in a trial (which is not the same as adherence with intended treatment, lack of efficacy, or the observation of adverse events) may present serious problems both in terms of generalizability of the findings to the wider population and, in the case of differential loss to follow-up across treatment groups, to the validity of the comparisons. Indeed, strictly speaking any missing outcome data means that not all of those allocated to the various randomization groups can be included in the analysis [86], and this might lead to the conclusion that the term 'intention-to-treat' should only be used if follow-up is complete. Under current guidelines, **intention-to-treat** relates more to the broad strategy adopted by the researchers for data analysis [39, 87]. Results should always be accompanied by a full and clear statement of how deviations from intended treatment and missing outcome measures have been handled in the analysis. The discussion should include how missing outcome data may have affected the conclusions [86]. Sensitivity analyses can be used to test the exclusion of, or assumptions about, missing values; practical examples of such analyses are becoming more common [88]. Another design strategy, modified intent to treat (MITT), is common in drug trials. It requires the participant to take at least a single dose of the study medication in order to be included in the analysis.

It is essential that a statistical analysis plan be developed for the trial prior to implementation. The CONSORT statement provides an outline of the various stages of data analysis for RCTs [32, 72]. Here we present the underlying concepts of data analysis at a particular follow-up time relative to randomization, and considers initially the simplest case of a clinical trial with two treatment groups. Multiple treatment groups will be covered briefly, but repeated measurements of outcomes and interim analyses involve considerably more complex methods of planning and analysis, for which expert help is essential [71, 89].

- **Stage of data analysis**

The first stage of data analysis is to address the representativeness of randomized subjects compared to the target population of eligible patients. The number of eligible patients who were and were not randomized should be provided, along with reasons for excluding potential study participants. This aspect of the findings of the trial will only prove useful if all eligible patients are considered: in practice there is a tendency for researchers to avoid approaching certain potentially eligible patients (selective pre-screening), for any of a wide variety of reasons, and this behavior may introduce bias in which subjects from the target population are included in the study. The presentation of this information is facilitated by use of the CONSORT flow diagram [48, 72]. Use of the CONSORT guidelines is associated with improved quality of reporting of trials generally [32]. Descriptive statistics should also be given of important characteristics of health care professionals approached for involvement in recruiting subjects to the trial, both for those taking part and those declining.

The second stage of data analysis is to compare the two groups at randomization (baseline) including demographic, prognostic, and outcome variables. A common error at this point is to rely on statistical testing for these comparisons [47, 50, 70]. If the randomization procedure has been performed correctly, then any statistically significant differences in baseline characteristics are likely due to chance. Statistical testing of this kind is not a test of the comparability of trial groups; rather, it is a test of the allocation procedure [47, 50, 70]. It may be seriously misleading, particularly if lack of a statistically significant difference for a given characteristic is taken to imply comparability. Baseline comparability is best assessed by simply obtaining descriptive statistics for the groups and making a judgment as to whether any observed differences are likely to be influential or not. If differences are likely to be influential, they should be considered in the analyses. Notable exceptions to this are baseline measures of the outcome variables, which should be considered in the analysis regardless of the situation at baseline, since removing variance in the outcome measure that is purely attributable to differences between individuals at baseline has potentially marked benefits in terms of precision and power [50]. Investigators should consider stratifying the randomization on any strongly prognostic variable (for reasons of efficiency rather than bias). Since there are practical limitations as to how many variables a trial can stratify for, a technique known as minimization may also be considered [46, 74]. Any variables stratified or minimized at randomization should be allowed for in the analysis [50]. In incontinence research, variables such as prior failure of therapy in drug and surgical studies or the degree of anatomic support in surgical or injectable trials or degree/amount of incontinence might be considered important enough to stratify, or to include in the analysis if unequally distributed.

The next stage of data analysis is to perform the primary (comparative) analyses for the outcome variables. First, though, it is essential to derive and report actual numeric data – even if simply in the form of descriptive statistics – rather than just reporting for instance a percentage change, even if the latter are relevant and provided as well. Graphs can be misleading, especially when subsections of the scales are magnified, and should be used to supplement or clarify the numerical data,

not to replace it. Primary outcomes should initially be analyzed by intention-to-treat comparisons of the groups as randomized, both using hypothesis tests for statistical significance and CIs for comparisons between the groups to assess clinical and statistical significance, usually adjusting for baseline measurements of the outcome variable. With a small number of primary outcomes, multiple testing is not a concern. However, when a large number of statistical tests are performed for secondary outcomes, corrections to the observed p-values should at least be considered.

The most commonly used procedure for multiple testing of many outcomes is the **Bonferroni correction** [47, 50, 82]. The Bonferroni correction is fairly conservative in reducing the risk of a statistically significant effect occurring purely by chance, at the cost of reduced power for individual outcomes. This is particularly pertinent when, as is usually the case, the outcomes are positively associated with one another. While there are alternative procedures that improve this deficiency, none of them are entirely satisfactory [50]. It is emphasized that whatever strategy is adopted to deal with multiple testing, the major errors are to rely solely on p-values rather than present confidence interals "CI"s as well, to over-simplify the presentation of p-values to just non-significant"NS" or "p<0.05" rather than to quote the actual p-values, and above all to report selectively the results of significance tests.

Another example of a **"multiplicity"** is where there are more than two treatment groups, e.g., when different doses of a drug are being investigated or when more than one 'active' procedure is being compared with placebo [50]. Similar issues to multiple testing of different outcomes are involved here, but there are a greater variety of commonly used procedures available to deal with the central concern of finding a difference purely by chance. Standard methods for dealing with this multiple comparisons problem include the procedures attributed to Tukey, Newman-Keuls and Dunnett [90].

Secondary analyses of trial data include per-protocol analyses with adjustments using regression methods for pertinent process measures such as degree of adherence with the allocated treatments. Secondary analyses also include **planned subgroup analyses,** such as the investigation of different intervention effects across age, ethnic, or disease severity groups. Subgroups should be analyzed by using appropriate interaction terms in regression models [48, 70]. Using interaction terms rather than performing repeated, separate, subgroup-specific analyses considerably reduces the risk of false positive findings [91-92]. Subgroup analyses should be carried out sparingly, specified in advance (preferably with a clinical rationale), and above all should not be reported selectively [48, 92-93]. This last point relates not just to subgroup analyses but to all stages of reporting randomized trials. Pre-specification of the primary outcomes in the study protocol and analysis plan, along with clear statements about all the outcomes considered is essential to avoid selective reporting. The large volumes of data accumulated in major multicenter RCTs almost guarantee that something "significant" can be identified by "data-mining". If not identified by the investigators as an a priori item of interest such findings should be viewed with great skepticism.

X. REPORTING RESEARCH RESULTS

Over that past decade, progress has been made in reporting research results. However, faster progress would occur if investigators would follow the reporting recommendations. As part of the "Enhancing the QUAlity and Transparency Of health Research" **(EQUATOR)** Network project, a website was developed at www.equator-network.org. On the website, the Consolidated Standards of Reporting Trials **(CONSORT)** statement provides guidelines for reporting the design, detailed methods, and results of RCTs. For studies of diagnostic tests the Standards for Reporting of Diagnostic Accuracy **(STARD)** [94] statement fills the same role and information is also included on the EQUATOR website. Guidelines for reporting meta-analyses are described by the Quality of Reporting of Meta-analyses (QUORUM-37) and Epidemiologic research reporting guidelines are contained in STROBE, Strengthen the Reporting of Observational Studies in Epidemiology [38, 72].

The CONSORT statement is specifically designed to provide standards for reporting RCTs [48, 72]. It includes a checklist of items that should be included by authors preparing manuscripts. Adherence to these guidelines and the use of flow diagrams in particular is associated with improved quality in reporting of RCTs [95]. Errors in presentation of statistical information are extensively covered in many textbooks [47, 82]. This section will emphasize the most important points on reporting of RCTs, to ensure an objective and comprehensive presentation of the trial itself, and also to facilitate any subsequent synthesis of research evidence including formal meta-analyses of RCTs. Meta-analyses are themselves the subject of separate reporting guidelines: the QUORUM statement [96]. However, such guidelines are not a panacea [97]; deficiencies in reporting are still common [72]. [98]

The CONSORT statement for reporting **parallel group randomized trials** have recently been updated Statements have also been issued for pragmatic trials [99-102] and noninferiority and equivalence trials. The CONSORT statement includes a checklist of items and a comprehensive set of characteristics of a clinical trial to assist investigators to write reports of their findings, journal editors and

reviewers in the review of manuscript submitted for publication and permit consumers of the literature to critically appraise articles. Clear statements about the objectives of the trial, intended study population, and planned comparisons. Subgroup or covariate analyses should be clearly specified and justified. The method of randomization should be stated, as should the unit of randomization; in most cases, this will be the individual participant but occasionally an aggregate group of subjects will be allocated jointly in a cluster randomized design [56]. Cluster randomized designs are also now the subject of separate reporting guidelines [103], and involve particular complications in terms of data analysis [104]. For all trials, specifications for the sample size calculation (primary outcomes, target differences, etc.) should be stated and justified. In addition, the precision actually obtained in a study must be presented. This requires confidence intervals as well as the observed p-values, at least for primary outcomes but preferably for all outcomes. The principal confidence intervals should be for comparisons between the groups, rather than for differences in the outcomes within the trial groups [47, 50]. Results should include a trial flow diagram, with numbers and reasons for the exclusion of eligible subjects, the number randomized, and subsequent losses to follow-up [32]. Protocol deviations should be described and explained [70]. Harms of the trial should be described for each treatment group. Finally, the discussion should include a brief summary of the trial's findings, possible explanations for the results, interpretation of the findings in light of the literature, limitations of the trial including internal and external validity, and the clinical and research implications of the study[48].

Recommendations on Study Conduct and Statistical Methods

- The role of quality RCTs as providing the strongest level of evidence in incontinence research should be fully acknowledged by researchers, journal reviewers, and editors. **HIGH**
- Careful attention to the planning and design of all research, especially RCTs, is of the utmost importance. **HIGH**
- Appropriate expertise in biostatistics and clinical trial design should be employed at the design phase of a RCT and thereafter on an ongoing basis. **HIGH**
- The design, conduct, analysis and presentation of RCTs must be fully in accordance with the CONSORT Statement. **HIGH**
- The design, conduct, analysis and presentation of observational studies should follow STROBE guidelines. **HIGH**
- The design, conduct, analysis and presentation of meta-analyses should follow QUORUM guidelines. **HIGH**
- Reporting studies of diagnostic tests, including urodynamics, should follow the STARD state¬ment guidelines. **HIGH**

XI. SPECIAL CONCERNS FOR SPECIFIC STUDIES

1. BEHAVIORAL AND PHYSIOTHERAPY TRIALS

a) Terminology

The lack of consistent terminology severely hampers our ability to build a body of literature about conservative interventions. The terms "behavioral therapy", "lifestyle intervention", "conservative treatment", "non-surgical treatment", "physiotherapy", "biofeedback", and pelvic floor muscle exercise" are often used interchangebly and incorrectly to describe both the same, and different interventions. While such therapies are discussed elsewhere in this consultation, we here advance this committee's opinion about appropriate terminology:

b) Behavioral interventions

According to Oxford Advanced American Dictionary the term behavioral is «the way someone behaves, especially towards other people», and behavioral science is the study of human behavior. We recommend that behavioral science be limited to studies that evaluate how people do or do not behave as desired.

c) Lifestyle interventions

Lifestyle interventions for UI are discussed elsewhere in this consultation and have traditionally included change in diet, intake of caffeine and carbonated soft drinks, fluid restriction, weight loss, smoking cessation and advice of increasing general physical activity level. Behavioral science can be used to understand how and why people change life-style to, for example, adhere to exercise and weight loss programs, but it should not be used as term to replace specific therapies such as physiotherapy or pelvic floor muscle training.

d) Physiotherapy

Physiotherapy refers to assessment, prevention and treatment given by an authorized physiotherapist (ICS physiotherapy committee www.icsoffice.org). It involves "using knowledge and skills unique to physiotherapists" and, "is the service only provided by, or under the direction and supervision of a physiotherapist" (WCPT 1999). This implies that the term physiotherapy should only be used in trials where the professional providing the intervention is a physiotherapist. We recommend describing the actual intervention instead of using the term physiotherapy: eg pelvic floor muscle training

with or without biofeedback, electrical stimulation, pelvic floor muscle training with vaginal cones or resistance device etc. This accurately describes the intervention and is neutral towards profession. Publications should report the actual profession of the interventionist (e.g., physiotherapist, general practitioner, urogynecologist, urologist, midwife, nurse, fitness instructor), rather than using the vague term, "therapist".

e) Biofeedback

Biofeedback encompasses "a group of experimental procedures where an external sensor is used to give an indication on bodily processes, usually in the purpose of changing the measured quality" [105]. Biofeedback equipment was developed within the area of psychology, mainly for measurement of sweating, heart rate and blood pressure under different forms of stress. Today, a variety of biofeedback apparatuses are commonly used in clinical practice to assist with PFMT, and the biofeedback can be either visual, auditory or both. In many textbooks the term "biofeedback" has been used to classify a method as different from PFMT. However, biofeedback is not a treatment on its own. It is an adjunct to training. For example, it may measure the response from a single PFM contraction or provide visual feedback during attempts to relax the muscles. Hence, a more precise terminology is PFM training or relaxation with or without biofeedback.

In addition to traditional biofeedback apparatuses, other instruments can offer valuable feedback. Vaginal and anal surface EMG, urethral, vaginal or anal squeeze pressure measurements, ultrasound and MRI can all be used to make the patients more aware of muscle function, and to enhance and motivate patients' effort during training [106].

f) Conservative interventions/treatments

Conservative interventions include all of the above, and this term includes everything except medication and surgery. As several studies have found that more than 30% of women with pelvic floor dysfunction are not able to perform a correct PFM contraction at the first consultation [107-110] it is mandatory to report how ability to perform a correct contraction was assessed before commencing an exercise trial. Also in electrical stimulation research one should report in which way response to the stimulation was assessed.

Although there is level A, grade 1 evidence for some of the conservative interventions such as PFMT for SUI, a variety of new conservative methods to treat the condition are frequently suggested in clinical practice and in the literature. These new methods are often presented as being effective, but are usually a hypothesis based on theories, data from small experimental laboratory studies/ small clinical series, or associations found in large epidemiological studies. If the clinicians like the approach, it soon enters clinical practice and textbooks without any further critical testing [113]. A model for how new therapies ideally should enter clinical practice has been proposed [114]. In this model the new idea should go through at least the 4 first stages before they enter practice:

Clinical observation and laboratory studies

- Clinical exploration
- Pilot studies
- Randomized controlled trials
- Refinement, eg dose-response issues
- Active dissemination of the method if it has proven to be effective

It is important that the patients participating in the 3 first stages before the RCT are given full information that they are participating in an experimental treatment, and that the clinician does not know if this new approach is effective. The patients also need to be informed if there are other proven effective treatments available.

g) Reporting of trial characteristics

In addition to reporting the specific type of intervention (eg.PFMT with biofeedback compared to PFMT alone or electrical stimulation) and the profession adminstering the intervention, the intervention needs to be described in such detail that other investigators can reproduce the intervention. This includes:

- Ability to perform a correct PFM contraction.
- Frequency: number of home training sessions and supervised training sessions (eg every day, 3 times/week)
- Number of repetitions and sets (eg. 12 sets x 3 times/ day)
- Duration: length of each training session (eg. 20 minutes), and duration of the total training period (eg, 3 months, 6 months)
- Intensity: In exercise science, this is usually reported as % of one repetition maximum for strength training. In pelvic floor muscle training it is often described as attempts to reach maximum contractions or utilizing submaximal contractions. Another description of intensity is the holding time in seconds, eg: 6-8 sec
- Adherence: the degree to which participants follow the prescribed protocol, usually reported as a percentage of the total possible.

Educational material provided, such as DVDs, brochures and booklets should be described.

Assessment: All devices used for assessments (eg manometers, dynamometers, ultrasound and EMG) must be described in detail, and their re-

sponsiveness (ability to detect small changes), reliability and validity should be reported [111].

h) Adherence vs effectiveness

It is important to note that adherence is not the same as the effectiveness of a program, as it is possible to have high adherence, but still little effect of training. Hence, when reporting the effect of conservative interventions it is ideal to measure also the exposure variable that the treatment is expected to change (eg muscle strength, ability to relax etc). This variable should not be confused with the primary or secondary outcomes of the intervention (eg leakage measured with pad testing, number of leakage episodes or QoL).

In many areas of conservative interventions there are high quality RCTs, systematic reviews and meta-analysis showing statistically significant and clinically relevant differences between the intervention and the untreated control group or other interventions. Of conservative therapies, PFMT for SUI/MUI has the strongest evidence to support its use; further, the more intensive the program (more supervision, higher dosage of training) the better the effect. Therefore, when comparing new methods and innovations with established PFMT, it is important to compare the new intervention with the current best evidence, meaning the effective arm of the reported RCTs. Unfortunately it is common to compare new methods with an ineffective training protocol, thereby overestimating the effect of the new method and claiming that it is equal to or better than “the old method”. When comparing different methods the dosage also needs to be the same in both treatment arms, eg when comparing PFMT with and without biofeedback, the number of supervised sessions, length of the sessions, frequency of home treatment and duration of the intervention must be the same.

i) Adverse events and cost

There are few adverse effects or complications reported after conservative interventions, but they do exist, eg in electrostimulation [112], and adverse effects or lack of adverse effect, and inconvenience to the patients should be reported. Although seldom harmful, conservative treatments are time consuming and can be costly for participants and paying parties because of the need for close follow up during the interventions. Cost effectiveness studies are crucial to fully understand where conservative therapies fit in the treatment armamentarium.

j) Outcome measures

The need for use of responsive, reliable and valid outcome measures in research is covered elsewhere in this chapter. The RCTs published in conservative treatment have applied a huge variety of outcome measures, making systematic reviews and meta-analysis difficult or impossible to conduct. Therefore, in future research it is important to use established and recommended outcome measures. In addition to description and use of responsive, reliable and valid primary and secondary outcome measures, future studies should include description and assessment of adherence to the intervention protocol, measurement of the independent variable (the intervention; eg strength training, relaxation training) and measurement of the possible underlining mechanisms of how the treatment works. It is usually not possible to blind the participants or those providing the intervention, but the assessors of outcome should always be blinded to group allocation.

k) Specific and non-specific effects

There have been some concerns that the effect of conservative treatments can be attributed to non-specific effects such as the extra attention of the therapist. The role of the therapist is to educate, motivate and empower the patient to be able to perform the actual program, secure high adherence, and minimize drop-outs. In patient reported outcomes and reports on quality of life it may be difficult to separate the effect of the attention and the actual effect. However, the effect of the attention is less likely to affect outcomes such as muscle strength, urodynamic assessments, pad testing and morphological changes measured by ultrasound and MRI. To minimize bias, all assessments should be conducted by investigators blinded to treatment allocation; the logistics of this should be addressed during the planning phase of the study. In a high quality RCT, Dumoulin et al address the problem of attention in physiotherapy research [115]. Women with persistent UI, three months after childbirth were randomized to either two different training regimens or a control group receiving relaxation massage for the back and limbs for the same amount of time as the supervised training groups. 70% were cured on pad testing in both treatment groups while there was no effect on urine loss in the relaxation massage group. Participants in the massage control group had improvement in disease specific quality of life.

l) Power calculations and number of participants

Some of the RCTs on conservative treatments are flawed by small sample sizes, this being especially evident in electrostimulation studies and may account for negative effects caused by type II errors. It is important that future studies use results of previous published studies to make appropriate power calculations that incorporate estimates of drop-outs and loss to follow-up to decide the optimal number of participants needed. Recruiting large numbers of participants may come at the expense of the rigor of the intervention [116]. Weak interventions (eg non-optimal training dosages or

suboptimal electrostimulation parameters) are unlikely to be effective and do not yield the true effect of an intervention. In meta-analyses, adding RCTs with large sample sizes but weak and ineffective interventions, can dilute the effect of smaller RCTs with higher methodological and interventional quality [116].

m) Long term studies

To date there are no quality criteria for how to report long term follow-up studies or how to conduct meta-analyses of long term studies. Challenges in long term follow-up studies include cross over to the more effective treatment after cessation of the original RCT, co-interventions during the follow-up period, recurrent events (eg new pregnancy), competing events (other diseases leading to incontinence) and loss to follow up. For conservative interventions it is expected that any training effect will diminish over time if no maintenance training is conducted or the pre- or co-contraction of the pelvic floor muscles has not reached an automatic level during the original trial. In order to control for as many of the above mentioned factors as possible, it is recommended that the long term follow up study should be planned together with the original RCT. Loss to follow-up and adherence to the protocol during the follow-up period must be reported [106, 111-117].

Recommendations for conservative treatment trials:

- Use correct terminology to describe the intervention. HIGH
- Report details of ability to perform correct contraction, dose-response issues and adherence. HIGH
- Use recommended outcome measures with high responsiveness, reliability and validity. HIGH
- Compare new methods with the best available intervention. HIGH
- Use power calculation in planning of the study. Avoid large sample sizes and weak (ineffective dosages) interventions. HIGH
- For long-term follow-up studies report crossover, co-interventions, recurrent and competing events, adherence in the follow-up period and loss to follow-up

2. EXPERIMENTAL DEVICES AND MATERIALS

Surgical research presents unique challenges to efforts at optimizing patient care. It is important to create a pathway for real advances while simultaneously protecting patient safety. When new procedures are substantially different from prior operations there should be a broad based preliminary exploration leading to a comparative trial if warranted. At the same time, many minor modifications of surgical procedures are inappropriate for randomized trials and if required, surgical progress would be slowed [118].

a) Randomisation

It has been argued that the first patient in whom a procedure is performed should be randomized [119-120]. Alternatively, it has been suggested that case series for new procedures are allowed until the procedure finds its intended use and to avoid doing studies while those performing the procedures are on the “learning curve”. Typically, new surgical procedures for incontinence have been reported as case series [121-122]. Not only do surgical case series provide the lowest level of evidence for treatment effects, case series may be “harmful”. An accumulation of “positive” case series may present a premature certainty about benefits of a procedure and make it even more difficult to perform randomized trials [123-124]. Influential members of the surgical community may endorse a new procedure and if the procedure is considered better it may be difficult to get surgeons and patients to randomize or a trial may appear to be unethical with a “proven” procedure [119, 123, 125].

b) Adoption by clinicians

Therefore, devices often are widely adopted by clinicians based on anecdotal data, marketing, or small case series. This raises a unique problem for trials in this area: 1) Surgeon buy-in can be difficult to obtain as some surgeons (“early adopters”) may perceive that the newest therapy is best and thus be unwilling to randomize patients to receive the traditional therapy, 2) Other surgeons (“late adopters”) may perceive that the data available do not support use of the newest therapy and thus be unwilling to randomize patients, 3) Patients may be unwilling to be randomized to traditional therapy because they are influenced by marketing forces propelling the newest devices to the forefront, and 4) Device companies frequently modify their materials or technique recommendations; therefore, by the end of the 3-5 years it typically takes to complete the earliest outcome assessment for a surgical randomized trial, the device tested is no longer the same as the device used in the trial. Thus, the results may be discounted as being no longer applicable.

c) Recruitment procedures

An important area of concern in surgical and device studies is patient recruitment procedures. Both in this regard and for all components of a study, we strongly support reporting according to the CONSORT guideline for randomized trials and the STROBE guidelines for observational studies. Subjects should be enrolled in a manner that minimizes selection bias. The protocol should detail the procedure by which consecutive patients meeting

the inclusion criteria are selected. All situations in which a patient meets the inclusion/exclusion criteria but is not offered enrollment by the investigator should be documented. The number of patients who decline enrollment should be stated, along with the reasons. It is vital that clinician researchers do not "cherry pick" from their patients, that is, that they do not limit recruitment to those patients considered to have the greatest chance of cure or lowest chance of risk. The study population should be as generalizable as possible. There should be a complete accounting of all participants in the study including the reasons for subject withdrawal.

Participants must be well informed about what is known and not known about devices or procedures being tested. They should not be led to assume that because a device is on the market, it is "safe and effective", as gaining knowledge about this is the purpose of the trial.

3. SURGICAL STUDIES

Cross-sectional studies of surgical procedures by type can provide estimates of prevalence, variation by age, race, and region as well as morbidity and mortality [126-127]. This type of information raises important health policy questions regarding physician practices, patient preferences for incontinence treatment, and differential access to and the utilization of care.

Case series are the most common study design found in the surgical literature, especially for new "innovative" surgical procedures. This is true despite the fact that case series cannot account for selection bias on the part of both the patient and surgeon, non-reporting bias of failures or loss to follow-up, lack of long-term follow-up, and provide the lowest level of evidence for treatment effects. Observational studies can provide important information about effectiveness and complications of surgical procedures, and also are very helpful in designing and selecting potential randomized clinical trials.

The randomized controlled trial is the accepted "gold standard" for research of treatment effects. In all surgical specialties, there has been growing concern regarding the limited number of randomized controlled trials for surgical procedures, poor methodological standards in those that have been performed, and a perception that surgeons are reluctant to rigorously test new surgical interventions [119, 123, 128-129]. A number of reasons for the paucity of surgical trials have been suggested including the lack of a regulatory board similar to the Food & Drug Administration responsible for the development of new medications [130].

The body of literature of surgical randomized trials in pelvic floor disorders is small, though increasing. However, many RCTs have serious limitations including a sample size too small to detect differences between groups, lack of blinding of the participants and/or individuals assessing the outcomes, short follow-up, inclusion of limited number of surgeons only, poor description of the technique, and lack of standardized outcome measures. Multi-center treatment networks are useful to overcome some of these limitations.

Differential drop out after randomization (or for cohort studies, after the intervention) can introduce bias. In a large and methodologically sound randomized controlled trial comparing the tension-free vaginal tape (TVT) and colposuspension, a large number of women withdrew from the colposuspension arm after randomization [131-132]. The loss of participants after randomization introduced bias in favor of the TVT because the drop outs had less severe incontinence resulting in the colposuspension group having more severe incontinence. It has been suggested that participants were only willing to continue if they were randomized to the "new and better" TVT procedure [132- 133]. Accounting for subjects "lost to follow-up" must also be detailed as per the CONSORT and STROBE recommendations. In the UK TVT RCT, drop out after surgery was similar for both procedures. In contrast, in two large study comparing two incontinence procedures performed by the Urinary Incontinence Treatment Network, randomization was done in the operating room after the patient was anesthetized and therefore no participant was lost to randomization [134],[67].

For studies of specific surgical procedures, the technique should be described in such detail that it could easily be reproduced in another study. **Standardization of the procedure** may vary depending on the research question [135]. Surgical trials using a small number of highly skilled surgeons are analogous to medical trials where only compliant patients are randomized, reflecting efficacy of the procedure in an ideal setting. It may be more generalizable to a mixture of skill level among surgeons in the community, and so reflect effectiveness of the procedure in usual practice [124].

Masking of participants as to their assigned intervention and those assessing the outcome is particularly important for surgical trials for incontinence because there may be enthusiasm by the patient or surgeon for a new procedure, many outcomes are based on the patient's own assessments such as symptom and quality of life scores, and the intervention is primarily for improvement of symptoms [118].

For new surgical procedures, important issues of adequate informed consent and conflicts associated with incentives for developing, starting and using new procedures have been raised. **Informed consent for a new procedure must include**:

- acknowledgement that the procedure is new and has not been shown to be more effective than a traditional approach
- discussion of potential complications, especially any integrally related to the procedure or device
- disclosure that information on complications are limited, and
- disclosure that the long-term benefits are unclear [122].

Incentives for adopting new procedures prior to sufficient evidence can arise from self-interest by attracting patients to one's practice, industry marketing, and patient desire for "cutting edge" techniques. Industry sponsorship or a surgeon's financial interest must be disclosed.

Organizations and treatment networks have been established to address many issues related to surgical interventions. Examples include the UK National Institute of Clinical Excellence (NICE www.nice.org.uk), the Australian Safety and Efficacy Register of New Interventional Procedures-Surgical (ASERNIP-S www.surgeons.org/asernip-s), and the US treatment networks: Urinary Incontinence Treatment Network (UITN http://www. niddk.nih.gov/patient/uitn/uitn.htm) for the NIDDK and the Pelvic Floor Disorders Network (PFDN). The NICE and ASERNIP-S provide systematic reviews of new operations, assessment of effectiveness, and recommendations that the technique has sufficient data for widespread use, or that the techniques appear unsafe, or that further audit/research are required before its widespread usage. The UITN and PFDN were established to provide the infrastructure for multicenter large randomized controlled trials for incontinence and prolapse.

Recommendations for Surgical and DeviceTrials:

- The safety and serious side effects of new operations must be completely defined with adequate follow-up so that risks can be weighed against efficacy. At a minimum, this requires more use of large scale, independent, prospective, multicenter cohort studies when RCTs are not practical. **HIGH**
- Safety and serious side effects of incontinence devices must be completely defined with adequate follow-up, especially for use of implantable devices and biologic materials, so that risks can be weighed against efficacy.**HIGH**
- Valid informed research consent is required in all trials of surgical interventions, which is separate from the consent to surgery. **HIGH**
- We recommend ongoing research into the usefulness of pre- and post-operative predictive testing (such as urodynamics, ultrasound, MRI, etc) in surgical trials. **HIGH**
- Reports of successful treatment should be limited to subjects with a minimum (not mean) of one year follow-up and should include a patient perspective measure. Specific assumptions about subjects lost to follow-up should be stated. **HIGH**
- Randomization for surgical trials should occur at the time of surgery to minimize dropouts and switch of procedure **HIGH**
- Long-term follow-up of RCT cohorts in an observational cohort is recommended. **HIGH**

4. PHARMACOTHERAPY TRIALS

Although many RCTs have been published in recent years on pharmacotherapy for urinary incontinence a great deal more remains to be learned. The trials have almost all been limited to 8-12 weeks of treatment giving very little information about long term safety and efficacy of drug therapy. Inclusion criteria are often stringent, such that the study population of healthy middle-aged people bears little resemblance to the patients for whom providers wish to prescribe medication. There is less than adequate information about special patient groups—men, children, neurogenic patients, and especially the frail elderly. Because incontinence creates such an impact on the older population, good studies to define the utility and safety of drug therapy are greatly needed in this group.

An issue of special relevance in trials of pharmaceutical agents (although germane to other treatment modalities) is the controversy regarding placebos in clinical trials. Regardless of whether a drug is effective or not, simply giving a drug to a patient may produce a beneficial response. To assess if a drug has an effect over and above the placebo response, it is usually tested against an inactive substance (placebo). In incontinence studies, the placebo effect may be quite large, anywhere from 30-50% in recent published studies. To account for this, investigators and regulators have generally demanded a placebo arm in most clinical trials of medication. While this may be acceptable to participants for short trials, it is neither ethical nor feasible to withhold treatment for longer periods of time. Further, clinicians and patients generally want to know how a new drug compares with established treatment.

Masking, while desired in all types of trials, is especially important in pharmacological trials. Further, it is feasible to do in such trials (as opposed to surgical or conservative interventions, in which masking may not always be possible) and thus should be prioritized. However, the identical appearance of two pills does not guarantee that participants will be unaware of group assignment. Side effects

common with anti-cholinergic therapy, such as dry mouth, may unmask participants. Studies should assess the degree to which masking was successfully maintained.

Recommendations for Pharmacotherapy Trials

- As effective drug therapy is available for most forms of incontinence, active drug comparator arms are recommended for most trials. **HIGH**
- Very little is known about the safety, efficacy and tolerability of drug therapy beyond 12 week trials. A concerted effort is needed to create this type of information base. Long-term follow-up of RCT cohorts in an observational cohort is recommended **HIGH**

XII. COST ANALYSIS

Economic and health policy outcomes are gaining increasing importance, as policy makers deliberate the values of different therapies. The financial burden on the health care system, the patient and patient's family of various treatment options makes cost an important outcome to measure. We recommend that cost analyses be planned with clinical studies whenever possible. Costs may be influenced by economic and political factors that are subject to change at any time; however, when basic units of work, time, and resources are carefully defined, models of costs remain useful even if market forces change in an unforeseen manner.

In health and medicine, economic analyses are descriptive and/or comparative. Descriptive data include the socioeconomic cost caused by the disease and its current treatment, whereas comparative data provide an economic evaluation of different treatment strategies and interventions where costs are compared to health outcomes.

There are several relevant types of cost analysis, some of which require a high level of expertise to conduct:

- **Cost of illness analysis** (COI) typically quantifies the burden of medical expenses (direct costs) and the resulting value of lost productivity (indirect costs) attributable to a specific condition such as an illness or injury [136], [132].
- **Cost effectiveness analysis** (CEA) measures the costs and consequences of two or more diagnostic or treatment pathways related to a single common effect or health outcome. It then summarizes the results in ratios that demonstrate the cost of achieving a unit of health effect for different types of patients and for variations of the intervention [137], [133].
- **Cost utility analysis** (CUA) is a form of cost effectiveness analysis in which particular attention is paid to the quality of health outcome related to treatment. In CUA, health effects are expressed in terms of quality-adjusted life years (QALYs) [138], [107]. A QALY is a measure of health outcome that assigns to a given period of time a weighting that corresponds to the health-related quality of life during that period, and then aggregates these weights across time periods. The QALY is important because it considers both quantity and quality of life.
- **Cost benefit analysis** estimates the net social benefit of an intervention by comparing the benefit of the intervention with the cost, with all benefits and costs measured in dollars [24], [134]. Health outcomes are converted into monetary values using "willingness to pay" (the value an individual would pay for reduction in illness severity) or "risk of death" or "human capital" methods (an individual's value to society based on productivity or future wages) [139- 140], [135-136].

Recommendations on Cost Analysis in Incontinence:

- Cost analysis should be incorporated into clinical studies whenever possible [137]. **HIGH**

XIII. RECOMMENDATIONS FOR SPECIFIC PATIENT GROUPS

1. MEN WITH LUTS

When considering men with LUTS, one must consider some unique factors which may influence urinary tract symptoms independently of any intervention, and so confound any data. These are the presence of the prostate gland which can cause bladder outlet obstruction (BOO), and the rarity of sphincter incompetence except in men who have undergone surgery for benign or malignant prostatic disease. For short term outcomes after intervention studies, these factors are unlikely to be relevant, but longer term follow up, and large observational or epidemiological studies may need to take these factors and changes over time into account when analyzing data. The prostate gland may complicate research outcomes as a result of outflow obstruction (either at baseline, or development of a new problem during follow up studies). Also, for patients with prostate cancer (either at the time of enrolment, or during follow up of longer studies), it is likely that both the disease, and the treatment given (surgery or radiotherapy) may alter urinary tract function and symptoms independently of any intervention in the study and thus confound the outcomes. , Overall, about 2/3 of men with LUTS have urethral obstruction and over 50 % have detrusor overactivity, al-

though a much smaller number have urinary incontinence due to detrusor overactivity [141].

If prostate size is considered to be a variable that could affect outcomes, measurement of prostate volume should be made before and after treatment. The method used to measure volume and its reliability and validity should be provided if available or their absence indicated. Any associations between outcomes and change in prostate size should be tested for using appropriate methods and reported. Consideration should be given to stratifying participants by prostate volume when there is suspicion that response to therapy may be size dependent.

Insofar as about 2/3 of men with LUTS have bladder outlet obstruction (BOO), any research protocol in men should consider inclusion of a method to screen for it. At the least, maximum free urinary flow rates and measurement of post-void residual urine should be recorded before and after treatment and the effect of therapy on these parameters should be documented simultaneously with assessment of the primary outcome variables. Synchronous pressure-flow studies are generally desirable and should be included whenever feasible. Several pressure-flow nomograms have been proposed to diagnose obstruction in men. The ICS nomogram is recommended and it is important to specify which if any nomogram is being used [142]

Recommendations for Research in Men:

- Measurement of prostate size should be performed before and after treatment (at the same time as continence outcome measures where possible) whenever prostate size is considered to be a potentially important variable, or to change during the intervention and follow up. **HIGH**
- Maximum free flow rate and measurement of post-void residual urine should be recorded pre-treatment and the effect of therapy on these parameters should be documented simultaneously with assessment of the primary outcome variables. **HIGH**
- Participants should be stratified by prostate size at randomization when size is considered to be a potentially important determinant of treatment outcome. **LOW**

2. WOMEN WITH LUTS

a) Hormonal effects

Our knowledge of hormonal influences on the lower tract remains limited. Recent RCTs and prospective cohort studies have demonstrated that hormone replacement therapy (HRT) does not improve or may worsen incontinence [143-145). It therefore seems appropriate that information about menstrual and hormonal status should be an integral part of the baseline history. New studies designed to examine the influence of hormones on incontinence (if considered ethical by an appropriate review board) should include details of hormonal status (premenopausal, postmenopausal without HRT, post-menopausal with HRT), the route and type of HRT (oestrogen only, combined sequential, combined continuous), and whether or not oophorectomy has been performed.

b) Obstetric History

The influence of vaginal childbirth on the structure and function of the female pelvis is the focus of much recent and ongoing research but the complex interactions remain incompletely understood. While it is clear that childbirth, and particularly vaginal childbirth increases the risk of incontinence and pelvic organ prolapse, the potential effect of further childbirth on previous or current treatments of incontinence (especially surgery) has yet to be determined.

Potentially confounding variables include: number and route of deliveries (vaginal/Cesarean), use of forceps or vacuum, infant birthweight and head circumference, duration of second stage of labor, use of episiotomy and any vaginal or perineal trauma, and epidural anaesthesia. The importance of these variables will depend upon the specific study design; for randomised studies the allocation process should balance these between groups, but consideration should be given to stratifying or minimizing the randomization against one or more of these important factors, depending on the exact intervention. For epidemiological or observational research each of these factors should be collected and included in univariate and multivariate analyses.

c) Pelvic Organ Prolapse

The effect of pelvic organ prolapse on lower urinary tract function remains poorly understood. Pelvic organ prolapse may potentially affect lower urinary tract function by obstruction of the urethral outflow, and thereby mask sphincter weakness (so called “occult” stress incontinence). Emerging evidence suggests that prolapse may contribute to the symptoms of OAB and correction of the prolapse may modulate these symptoms. Thus, it is important to include assessment of pelvic organ prolapse in incontinence research on women. A validated assessment method for prolapse should be used to identify the stage of proplase; the Pelvic Organ Prolapse Quantification System (POP-Q) [9] is recommended. Research protocols should be developed, either excluding women with prolapse severity beyond a specified stage, or the analysis plan should include stratification for stage of prolapse in randomization, and adjustment for prolapse stage in any analysis. For larger studies where regression analyses are planned, stage of prolapse should be considered a mandatory factor for inclusion. Prolapse should be graded at the same time as the outcome assessment for incontinence and LUTS is performed.

Recommendations for research in women:

- Specific information about menopausal status, hysterectomy, parity/obstetric history, and hormonal status should be included in baseline clinical trial data and controlled for in specified analyses in the research protocol. **HIGH**
- High quality, symptom and bother scores (e.g.,ICIQ-FLUTS, KHQ, PISQ, ICIQ-FLUTSsex) validated in women should be employed when assessing outcomes **HIGH**
- Standardized assessment of pelvic organ prolapse (by POPQ) should be performed before treatment and at the time of other outcome assessments in all research where prolapse and continence outcomes are being assessed. **HIGH**
- Criteria for cure/improvement/failure from incontinence treatment should be defined in the protocol based on patient perception as well as objective and semi-objective instruments such as validated questionnaires, diaries and pad tests. **HIGH**
- Assessment of the impact of treatment on sexual function should be performed with other outcome assessment when appropriate. **MEDIUM**

3. FRAIL OLDER AND DISABLED PEOPLE

There are a number of unique and pertinent research issues for this population.

In the frail elderly, important variables include:

- **Demographic information:** Advancing age, white race, and women [145-148) are associated with an increase risk of incontinence and each of these variables should be adjusted for in most analyses.
- **Medical Conditions:** Medical conditions related and unrelated to the lower urinary tract have been shown to increase the risk of incontinence in older women and are especially important to assess in the frail older population [147, 149-151]. Prior hysterectomy has also been suggested as a potential risk factor for incontinence in older women [152-153].
- **Medication Inventory:** Certain medications may exacerbate incontinence and therefore a complete medication inventory is essential [151, 153-155].
- **Physical function:** Mobility is often impaired in the frail elderly and impacts urinary control [156], therefore mobility should be assessed using validated instruments such as the Bartel Orcats or ADL scales [156-157]. Data on walking aids or wheelchairs, gait speed, and manual dexterity may also be collected.
- **Cognitive function:** Cognitive function impairment and/or dementia increase the risk of incontinence [158]. The Mini-Mental Status Scale Examination [130] assesses global cognitive function, and the Confusion Assessment Method (CAM) [159] is a standardized assessment for delirium. A battery of neuropsychological tests to measure subtle impairments in cognitive function include the Buschke Selective Reminding Test (verbal learning and memory) [160], the Digit Symbol (incidental memory, visual scanning and motor speed) [161], and the Trails A (attention and visual) [162].

Outcome measures should be selected for applicability to the frail elderly. Commonly used self-reported measures of frequency of urinary symptoms, severity, or level of bother may not be possible in the cognitively impaired frail elderly patient. Similarly, voiding diaries that have been shown to be valid and reliable in assessing urinary frequency, nocturia, and incontinence episodes by type [138, 162-164] may not be feasible or reliable. Motivated and trained staff, caregivers, or family members may be able to adequately collect diary data; however, this has not been validated.

In nursing home or inpatient settings, **wet checks by staff** at set intervals have been used in a number of studies. There are limitations to the measurement including visually determining what is “wet” because of new absorbent materials and staff reports not always being reliable or valid, due to underreporting [165-166]. To overcome the limitation of defining wetness and underreporting, 24-hour **pad weighing tests** [167-168] may be used. Pad weighing tests and wet checks are feasible and can provide important outcome data if staff is well trained and checks are often [14]. New outcome measures specific to the frail older population such as increased socialization or decreased caregiver burden need to be developed.

4. CHILDREN

The conduct of clinical research in children is generally more difficult than in adults. Four overriding issues separate pediatric research from the general recommendations. First, physiology varies widely within the group referred to as “children”, differs from adults, and changes with time. Because children are growing, any treatment, especially pharmacological and surgical therapy, may affect them profoundly in the long term. This is particularly true of the immature brain, nervous system and other incompletely developed systems. Second, compliance with therapy is more complicated as children may depend on caregivers to administer treatment in many studies. Third, reporting of symptoms and outcomes may be difficult. Symptoms reported by a caregiver may not be interpreted in the same way as the child. Finally, the issue of informed consent is complex with children.

Urinary incontinence in children falls into four main categories: neurogenic (myelomeningocele and other less common neurogenic etiologies), **monosymptomatic nocturnal enuresis, detrusor overactivity,** and **dysfunctional voiding without neurologic disease.** This issue of age groups is most crucial in children with myelomeningocele. These children are often on medication beginning at a very young age and continuing for many years; the long-term safety of medications in children must be established in all age groups. Therapy for other causes of incontinence in children tends to start at a later age, by which time size is the main difference between children. We recommend that clinical studies have long-term (five years or more), open label extension arms to monitor safety, particularly focusing on normal growth and development and the effects on treatment of liver and central nervous system function.

Assessment of compliance with therapy is always difficult, and even more so with children. Compliance with voiding diaries, a significant issue in the adult population, may be even more problematic with children. The social and family interactions between the child and parent or carer can influence the accuracy of data collection and treatment compliance, both positively and negatively.

Outcome measures are not as well developed in children as in adults. Validated, age-specific symptom and disease-specific quality of life instruments must be developed for the pediatric population. Early efforts in this area have been reported for dysfunctional voiding [169] and daytime incontinence [170]; much more work remains to be done. Invasive urodynamics can be used when deemed appropriate (specifically in the neurogenic population); however, the test-retest reproducibility of urodynamic investigations in children is still under investigation.

Recommendations for Research in Children:

- Long-term follow-up is of critical importance in the pediatric population in order to ascertain the effect of a treatment on normal growth and development. **HIGH**
- Research is needed to develop standardized outcome measures including validated, age-specific symptom and disease-specific quality of life outcome measures. **MEDIUM**

5. NEUROGENIC POPULATIONS

a) Classification

Classification of NLUTD has three primary aims—to aid in discriminating or identifying an unknown underlying neurological disease process, to characterize the nature of the dysfunction so as to develop a treatment plan, and to assess the risk of secondary effects (e.g. on the upper tract) which may influence the necessity and aggressiveness of treatment. The latter two are clearly relevant to research in neurogenic incontinence and must be reflected in study design and patient description.

It is difficult to find a classification system of NLUTD as a base for research that is satisfactory for each of the three aims. The published systems have been reviewed in detail [171]. Both the disease process and the site of the neurologic lesion(s) are relevant in the study of NLUTD, yet even this information is inadequate to predict the functional characteristics for an individual patient. There is no one method that meets the broad needs of classification in this group. Typical or classic cases are often well described but it is especially difficult to describe mixed and incomplete lesions. Thus, classification systems necessarily oversimplify or become extremely cumbersome. Finally, it must be acknowledged that the complexity of neurologic diseases and variations in individual behavior almost always call for a customized approach to therapy, further complicating research in the neurogenic patient. All of these factors complicate study design as it becomes difficult to create workable inclusion and exclusion criteria that apply to other than a narrow segment of the neurogenic population. Ideally a broad population of potentially relevant participants would be enrolled in research studies with full characterization of both the neurologic condition and the nature of the lower urinary tract dysfunction so as to allow for subgroup analysis.

b) History and evaluation:

Study planning is best undertaken with the cooperation of urologist, neurologist, and other clinicians, who have a specific interest and special training in the neurogenic patient. Baseline data collected by history in subjects with neurogenic lower urinary tract disorders should include:

- bladder volumes by diary or examination (maximum voided or catheterized volume, post voiding residual urine, total capacity);
- mechanism of bladder evacuation: normal (volitional), spontaneous involuntarily (“reflex”), Credé, sterile intermittent catheter (SIC), clean intermittent catheter (CIC), intermittent catheter by second person, or suprapubic or urethral catheter;
- use of external appliances (e.g., diaper or pad use, condom catheter, urethral catheter, suprapubic tube);
- the typical time span of continence (continence interval) following last bladder evacuation and maximal continent bladder volume.
- bowel function, sexual function, and specific neurologic deficits
- the evolution of the condition of (changes in)

the upper tract should be included in the outcomes evaluation of treatment for NLUTD.

Where possible these factors should be controlled for in analyses, and stratified for in randomization for interventional studies, or patients with certain factors should be excluded. The details will depend on the research question to be asked.

c) Urodynamics:

Baseline urodynamics are recommended for research studies of neurogenic incontinence, because the symptomatic response is usually of lesser relevance, compared to objective response in NLUTD, especcially where NDO or reduced compliance is involved in the dysfunction. Neurogenic disorders commonly cause complex and generalized lower tract dysfunction, often with combined bladder and urethral sphincter abnormalities.

Recommendations for Research in Populations Affected by Neurogenic Lower Urinary Tract Dysfunction:

- Detailed urodynamic studies are recommended for classification of neurogenic lower urinary tract disorders in research studies because the nature of the lower tract dysfunction cannot be accurately predicted from clinical data. Videourodynamic studies are preferred but are not mandatory. **LOW**
- An area of high priority for research is the development of a classification system to define neurogenic lower urinary tract disorders. Relevant features could include the underlying diagnosis, the symptoms, more precise documentation of the neuromuscular lesion by clinical neurophysiologic testing, and the nature of the urodynamic abnormality. **LOW**

6. RECOMMENDATIONS FOR RESEARCH IN POPULATIONS AFFECTED BY FAECAL INCONTINENCE

- High quality, validated, symptom and bother scores (e.g., ICIQ-BS, Wexner score, FQIL, Manchester Questionnaire, FISI) should be employed when assessing outcomes **HIGH**
- Due to the high concordance of faecal and urinary incontinence, and the potential for urinary incontinence therapy to affect bowel function, data on faecal incontinence should be collected at the outset and during trials of urinary incontinence whenever practical. **HIGH**

7. RECOMMENDATIONS FOR RESEARCH IN POPULATIONS AFFECTED BY BLADDER PAIN SYNDROME (INCLUDING INTERSTITIAL CYSTITIS)

- Broader entry criteria should be used to reflect the full spectrum of the BPS/IC patient population. **MEDIUM**
- The primary endpoint of BPS/IC trials should be patient driven and the Global Response Assessment is recommended. A wide spectrum of secondary endpoints will be useful in defining the effect of treatments. **MEDIUM**

8. RECOMMENDATIONS FOR RESEARCH IN POPULATIONS AFFECTED BY PELVIC ORGAN PROLAPSE

- A validated standardized assessment of prolapse (eg POP-Q) should be used for baseline and outcome assessments. **HIGH**
- Complete reporting of outcomes including a validated assessment of anatomy, functional status, and complications is essential. **HIGH**
- Complications/adverse events (especially for mesh) must be explicitly and completely
- Long term outcomes (> 2 years) of intervention studies are needed **HIGH**

XIV. CONCLUSIONS

The 2012 Consultation examined and classified available data in order to determine the level of evidence that supports our care of incontinent patients. This committee's contribution was to provide guidance to facilitate high quality research for the next Consultation. All quality research, be it prospective or retrospective, clinical or preclinical, begins with a clear research question and benefits from detailed planning—establishing a clear and relevant hypothesis, developing a trial of appropriate magnitude to accept or reject the hypothesis, and defining methods of adequate sensitivity and specificity to produce credible data.

REFERENCES

1. Abrams P, Blaivas JG, Stanton SL, Andersen JT. The standardisation of terminology of lower urinary tract function. the international continence society committee on standardisation of terminology. Scand J Urol Nephrol Suppl. 1988;114:5-19.
2. Bates P, Bradley WE, Glen E, Melchior H, Rowan D, Sterling A, et al. The standardization of terminology of lower urinary tract function. Eur Urol. 1976;2(6):274-6.
3. Bates P, Glen E, Griffiths D, Melchior H, Rowan D, Sterling A, et al. Second report on the standardization of terminology of lower urinary tract function. procedures related to the evaluation of micturition: Flow rate, pressure measurement, symbols. Act Urol Jpn. 1977;27:1563-6.
4. Bates P, Bradley WE, Glen E, Griffiths D, Melchior H, Rowan D, et al. Third report on the standardisation of terminology of lower urinary tract function. procedures related to the evaluation of micturition: Pressure flow relationships, residual urine. Euro Urol. 1980;6:171-.
5. Fourth report on the standardisation of terminology of lower urinary tract function. terminology related to neuromuscular dysfunction of the lower urinary tract. produced by the international continence society. Br J Urol. 1981 Aug;53(4):333-5.
6. Abrams P, Blaivas JG, Stanton SL, Andersen J, Fowler CJ, Gerstenberg T, et al. Sixth report on the standardisation of terminology of lower urinary tract function. procedures related to neurophysiological investigations: Electromyography, nerve conduction studies, reflex latencies, evoked potentials and sensory testing. Br J Urol. 1987;59(4):300-4.
7. Rowan D, James ED, Kramer AE, Sterling AM, Suhel PF. Urodynamic equipment: Technical aspects. produced by the international continence society working party on urodynamic equipment. J Med Eng Technol. 1987 Mar-Apr;11(2):57-64.
8. Andersen J, Blaivas J, Cardozo L, Thuroff J. Lower urinary tract rehabilitation techniques: Seventh report on the standardization of terminology of lower urinary tract function. Int Urogynecol J. 1992;3:75 - 80.
9. Bump R, Mattiasson A, Bo K, Klarskov P, Smith ARB, Brubaker L, et al. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. Am J Obset Gynecol. 1996;175:10-7.
10. Griffiths D, Hofner K, van Mastrigt R, Rollema HJ, Spangberg A, Gleason D. Standardization of terminology of lower urinary tract function: Pressure-flow studies of voiding, urethral resistance, and urethral obstruction. international continence society subcommittee on standardization of terminology of pressure-flow studies. Neurourol Urodyn. 1997;16(1):1-18.
11. Mattiasson A, Djurhuus JC, Fonda D, Lose G, Nordling J, Stohrer M. Standardization of outcome studies in patients with lower urinary tract dysfunction: A report on general principles from the standardisation committee of the international continence society. Neurourol Urodyn. 1998;17(3):249-53.
12. Lose G, Fantl JA, Victor A, Walter S, Wells TL, Wyman J, et al. Outcome measures for research in adult women with symptoms of lower urinary tract dysfunction. Neurourol Urodyn. 1998;17(3):255-62.
13. Nordling J, Abrams P, Ameda K, Andersen JT, Donovan J, Griffiths D, et al. Outcome measures for research in treatment of adult males with symptoms of lower urinary tract dysfunction. Neurourol Urodyn. 1998;17(3):263-71.
14. Fonda D, Resnick NM, Colling J, Burgio K, Ouslander JG, Norton C, et al. Outcome measures for research of lower urinary tract dysfunction in frail older people. Neurourol Urodyn. 1998;17(3):273-81.
15. Tjandra J, Chan M, Yeh C, Murray-Green C. Sacral nerve stimulation is more effective than optimal medical therapy for severe fecal incontinence: A randomized, controlled study. Dis Colon Rectum. 2008;51:494-502.
16. Blaivas JG, Appell RA, Fantl JA, Leach G, McGuire EJ, Resnick NM, et al. Standards of efficacy for evaluation of treatment outcomes in urinary incontinence: Recommendations of the urodynamic society. Neurourol Urodyn. 1997;16(3):145-7.
17. Blaivas JG, Appell RA, Fantl JA, Leach G, McGuire EJ, Resnick NM, et al. Definition and classification of urinary incontinence: Recommendations of the urodynamic society. Neurourol Urodyn. 1997;16(3):149-51.
18. Blaivas JG. Outcome measures for urinary incontinence. Urology. 1998 Feb;51(2A Suppl):11-9.
19. Protection of human subjects [Internet].: Department of Health and Human Services; 2005. Available from: http://www.hhs.gov/ohrp/.
20. Federman D, Hanna K, Rodriguez LL. Responsible research: A systems approach to protecting research participants. Washington, DC: The National Academices Press; 2002.
21. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000 May 24-31;283(20):2701-11.
22. Sung NS, Crowley WF,Jr, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA. 2003 Mar 12;289(10):1278-87.
23. Barnbaum DR, Byron M. Research ethics: Text readings. Upper Saddle River, NJ: Prentice Hall; 2001.
24. Gallin J. Principles and practices of clinical research. 2002: Academic Press; 2002.
25. Shamoo AE, Resnik DB. Responsible conduct of research. New York, NY: Oxford University Press; 2003.
26. Weijer C, Miller PB. When are research risks reasonable in relation to anticipated benefits? Nat Med. 2004 Jun;10(6):570-3.
27. Ethical priciples and guidelines for the protection of human subjects of research [Internet].: The Belmont Report; 1979. Available from: http://ohsr.od.nih.gov/guidelines/belmont.html
28. Donovan J, Mills N, Smith M, Brindle L, Jacoby A, Peters T, et al. Quality improvement report: Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. commentary: Presenting unbiased information to patients can be difficult. BMJ. 2002 Oct 5;325(7367):766-70.
29. Featherstone K, Donovan JL. Random allocation or allocation at random? patients' perspectives of participation in a randomised controlled trial. BMJ. 1998 Oct 31;317(7167):1177-80.
30. Hulley ST, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research. 3rd ed. Philadelphia: Lippincott, Williams & Wilkins; 2007.
31. Laine C, Horton R, DeAngelis CD, Drazen JM, Frizelle FA, Godlee F, et al. Clinical trial registration — looking back and moving ahead. N Engl J Med. 2007 06/28; 2012/02;356(26):2734-6.
32. Statistical considerations in the design of clinical trials [Internet].: International Conference on Harmonisation; May 29, 2001. Available from: http://www.ifpma.org/pdfifpma/e9.pdf.
33. Friedman LM, Furberg CD, DeMets DL. Fundalmentals of clinical trials. 3rd ed. New York: Springer Verlag; 1998.
34. Pagano M, Gauvreau K. Principles of biostatistics. 2nd ed. Pacific Grove, CA: Duxbury Press; 2000.

35. Vittinghoff E, Glidden DV, Shiboski SC. Regression methods in biostatistics. Springer; 2005.

36. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. N Engl J Med. 2011 Mar 3;364(9):852-60.

37. Hicks LK, Laupacis A, Slutsky AS. A primer on data safety monitoring boards: Mission, methods, and controversies. Intensive Care Med. 2007 Oct;33(10):1815-8.

38. Slutsky AS, Lavery JV. Data safety and monitoring boards. N Engl J Med. 2004 Mar 11;350(11):1143-7.

39. Grant A, et al. Issues in data monitoring and interim analysis of trials. Health Technol. Vol. 9(7) DOI 10.3310/hta9070

40. Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ, DAMOCLES Group. Reported use of data monitoring committees in the main published reports of randomized controlled trials: A cross-sectional study. Clin Trials. 2004 Feb;1(1):48-59.

41. Society for Clinical Data Management. Good clinical data management practices. 2nd ed. Hillsborough, NJ: ; 2002.

42. Lundberg GD, Glass RM. What does authorship mean in a peer-reviewed medical journal? JAMA. 1996 Jul 3;276(1):75.

43. Task force on financial conflicts of interest in clinical research protecting subjects, preserving trust, promoting progress II: Principles and recommendations for oversight of an institutions's financial interests in human subjects research [Internet].; 2002.

44. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. JAMA. 2003 Jan 22-29;289(4):454-65.

45. Spilker B. Guide to clinical studies and developing protocols. New York: Raven Press; 1984.

46. Pocock S. Clinical trials: A practical approach. Chichester: Wiley; 1983.

47. Altman D. Practical statistics for medical research. London: Chapman & Hall; 1991.

48. Simon R, Thall P. Phase II trials. In: Armitage P, Colton T, editors. Encyclopedia of Biostatistics. 1st ed. Chichester, UK: Wiley; 1998. p. 3370-6.

49. A proposal for structured reporting of randomized controlled trials. the standards of reporting trials group. JAMA. 1994 Dec 28;272(24):1926-31.

50. Senn S. Statistical issues in drug development. Chircheste: Wiley; 1997.

51. Peters T, Wildschut H, Weiner C. Epidemiologic considerations in screening. In: Wildschut H, Weiner CP, Peters TJ, editors. When to screen in obstetrics and gynecology. London: WB Saunders; 1996.

52. Rosenberger WF, Lachin JM. Randomization in clinical trials: Theory and practice. Colton T, editor. New York: Wiley; 2002.

53. Armitage P. Statisticals methods in medical research. Oxford: Blackwell Science; 1994.

54. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: A systematic review. JAMA. 2003 May 21;289(19):2545-53.

55. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol. 2003 Nov 24;3:26.

56. Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.

57. Fleming TR, Odem-Davis K, Rothmann MD, Li Shen Y. Some essential considerations in the design and conduct of non-inferiority trials. Clin Trials. 2011 08;8(4):432-9.

58. Schumi J, Wittes J. Through the looking glass: Understanding non-inferiority. Trials. 2011;12:106.

59. Kaul SD,George A. Good enough: A primer on the analysis and interpretation of noninferiority trials. Ann Intern Med. 2006 July 4;145(1):62-9.

60. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. Guidance for industry non-inferiority clinical trials. Silver Spring, Maryland: U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research; 2010.

61. Committee for Medicinal Products for Human Use, editor. Guideline on the choice of the non-inferiority margin. London, England: Eruopean Medicines Agency; 2005.

62. European Medicines Agency, editor. Choice of control group in clinical trials. London, England: ; 2001.

63. Nager CW, Brubaker L, Daneshgari F, Litman HJ, Dandreo KJ, Sirls L, et al. Design of the value of urodynamic evaluation (ValUE) trial: A non-inferiority randomized trial of preoperative urodynamic investigations. Contemporary Clinical Trials. 2009 Nov;30(6):531-9.

64. Siegel JP. Equivalence and noninferiority trials. Am Heart J. 2000 Apr;139(4):S166-70.

65. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. BMJ. 1996 Jul 6;313(7048):36-9.

66. Powers JH. Noninferiority and equivalence trials: Deciphering 'similarity' of medical interventions. Stat Med. 2008 Feb 10;27(3):343-52.

67. Richter HE, Albo ME, Zyczynski HM, Kenton K, Norton PA, Sirls LT, et al. Retropubic versus transobturator midurethral slings for stress incontinence. N Engl J Med. 2010 Jun 3;362(22):2066-76.

68. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. J Chronic Dis. 1967;20(637):648.

69. Newell DJ. Intention-to-treat analysis: Implications for quantitative and qualitative research. Int J Epidemiol. 1992 Oct;21(5):837-41.

70. Matthews J. An introduction to randomized controlled clinical trials. 1st ed. London: Arnold; 2000.

71. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd ed. Springer Verlag; 1998.

72. Moher D, Jones A, Lepage L, CONSORT Group (Consolitdated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. JAMA. 2001 Apr 18;285(15):1992-5.

73. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Ann Intern Med. 2002 Feb 5;136(3):254-9.

74. Egger M, Juni P, Bartlett C, CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. JAMA. 2001 Apr 18;285(15):1996-9.

75. Hilton P, Robinson D. Defining cure. Neurourol Urodyn. 2011;30(5):741-745.

76. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2(14):i,iv, 1-74.

77. Sung VW, Marques F, Rogers RR, Williams DA, Myers

DL, Clark MA. Content validation of the patient-reported outcomes measurement information system (PROMIS) framework in women with urinary incontinence. Neurourol Urodyn. 2011;30(4):503-9.

78. Dindo D, Demartines N, Claiven PA. Classification of surgical complications. A new proposal with evlaution in a cohort of 6336 patients and results of a survey. Ann Surg. 2004;240(2):205-13.

79. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. BMJ. 1995 Oct 28;311(7013):1145-8.

80. Machin D. Sample size tables for clinical studies. 2nd ed. Oxford: Blackwell Science; 1997.

81. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ. 2001 Jan 27;322(7280):226-31.

82. Bland M. An introduction to medical statistics. 2nd ed. Oxford: Oxford University Press; 1995.

83. Sterne J. Commentary: Null points--has interpretation of significance tests improved? Int J Epidemiol. 2003 Oct;32(5):693-4.

84. Kirkwood B, Sterne J. Essential medical statistics. 2nd ed. Oxford: Blackwell Science; 2003.

85. Hoenig JM, Heisey DM. Null points - has interpretation of significance tests improved. J Am Stat. 2001;55:19-24.

86. Hollis S, Campbell F. What is meant by intention to treat analysis? survey of published randomised controlled trials. BMJ. 1999 Sep 11;319(7211):670-4.

87. Lewis JA, Machin D. Intention to treat--who should use ITT? Br J Cancer. 1993 Oct;68(4):647-50.

88. Richards SH, Bankhead C, Peters TJ, Austoker J, Hobbs FD, Brown J, et al. Cluster randomised controlled trial comparing the effectiveness and cost-effectiveness of two primary care interventions aimed at improving attendance for breast screening. J Med Screen. 2001;8(2):91-8.

89. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ. 1990 Jan 27;300(6719):230-5.

90. Zar J. Biostatistical analysis. 2nd ed. New Jersey: Prentic Hall; 1984.

91. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol. 2004 Mar;57(3):229-36.

92. Brookes S. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol;In press.

93. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: Quantifying the risks of false-positives and false-negatives. Health Technol Assess. 2001;5(33):1-56.

94. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Fam Pract. 2004 Feb;21(1):4-10.

95. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. 2001 Apr 14;357(9263):1191-4.

96. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. quality of reporting of meta-analyses. Lancet. 1999 11/27;354(9193):1896-900.

97. Raz S, Erickson DR. SEAPI QMM incontinence classification system. Neurourol Urodyn. 1992;11(3):187-99.

98. Cavadas V, Frederico B, Carvalho F, Osorio L, Gomes M, Silva-Ramos M. The quality of reporting of randomized controlled tirals in pelvic organ prolapse. Int Urogynecol J. 2011;22:1117-11125.

99. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. BMJ. 2008 Nov 11;337:a2390.

100. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. JAMA. 2006 Mar 8;295(10):1152-60.

101. Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: Extension to cluster randomised trials. BMJ. 2004 Mar 20;328(7441):702-8.

102. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. Ann Intern Med. 2008 Feb 19;148(4):295-309.

103. Altman DG. Better reporting of randomised controlled trials: The CONSORT statement. BMJ. 1996 Sep 7;313(7057):570-1.

104. Peters TJ, Richards SH, Bankhead CR, Ades AE, Sterne JA. Comparison of methods for analysing cluster randomized trials: An example involving a factorial design. Int J Epidemiol. 2003 Oct;32(5):840-6.

105. Schwartz G, Beatty J, editors. Biofeedbak: Theory and research. New York: Academic Press; 1977.

106. Bo K. Pelvic floor muscle training for stress urinary incontinence. In: Bo K, Berghumans B, Morkved S, Van Kampen M, editors. Evidence-based physical therapy for the pelvic floor. Philadelphia, Pa: Churchill Lingstone Elsevier; 2007. p. 171-87.

107. Kegel AH. Stress incontinence and genital relaxation; a nonsurgical method of increasing the tone of sphincters and their supporting structures. . 1952;Ciba Clin Symp(4):35-51.

108. Benvenuti F, Caputo GM, Bandinelli S, Mayer F, Biagini C, Sommavilla A. Reeducative treatment of female genuine stress incontinence. Am J Phys Med. 1987;66(4):155-68.

109. Bump RC, Hurt WG, Fantl JA, Wyman JF. Assessment of kegel pelvic muscle exercise performance after brief verbal instruction. American Journal of Obstetrics & Gynecology. 1991;165(2):322,7; discussion 327-9.

110. Larsen S, Oseid S. Knowledge about and ability to correct pelvic floor muscle exercises in women with urinary stress incontinence. Neurourol Urodynam. 1988;7(3):261-2.

111. Bo K, Sherburn M. Evaluation of female pelvic-floor muscle function and strength. Phys Ther. 2005 Mar;85(3):269-82.

112. Indrekvam S, Hunskaar S. Side effects, feasibility, and adherence to treatment during home-managed electrical stimulation for urinary incontinence: A norwegian national cohort of 3,198 women. Neurourology & Urodynamics. 2002;21(6):546-52.

113. McKinlay JB. From "promising report" to "standard procedure": Seven stages in the career of a medical innovation. Milbank Mem Fund Q Health Soc. 1981 Summer;59(3):374-411.

114. Bo K, Herbert RD. When and how should new therapies become routine clinical practice? Physiotherapy. 2009 Mar;95(1):51-7.

115. Dumoulin C, Lemieux MC, Bourbonnais D, Gravel D, Bravo G, Morin M. Physiotherapy for persistent postnatal stress urinary incontinence: A randomized controlled trial. Obstet Gynecol. 2004 Sep;104(3):504-10.

116. Herbert RD, Bo K. Analysis of quality of interventions in systematic reviews. BMJ. 2005 Sep 3;331(7515):507-9.

117. World Confederation for Physical Therapy, editor. Description of physical therapy. London: ; 1999.

118. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: Problems and possible solutions. BMJ. 2002 Jun 15;324(7351):1448-51.

119. McLeod RS, Wright JG, Solomon MJ, Hu X, Walters BC, Lossing A. Randomized controlled trials in surgery: Issues and problems. Surgery. 1996 May;119(5):483-6.

120. Chalmers TC. Randomization of the first patient. Med Clin North Am. 1975 Jul;59(4):1035-8.

121. Nygaard I, Holcomb R. Reproducibility of the seven-day voiding diary in women with stress urinary incontinence. International Urogynecology Journal. 2000;11(1):15-7.

122. Gates E. Ethical considerations in the incorporation of new technologies into gynecologic practice. Clinical Obstetrics & Gynecology. 2000 September;43(3):540-50.

123. Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. Surgery. 1996 Apr;119(4):466-72.

124. Baum M. Reflections on randomised controlled trials in surgery. Lancet. 1999 Apr;353 Suppl 1:SI6-8.

125. Frader J, Caniano D, editors. Research and innovation in surgery. Houston, Tx: Oxford University Press; 1998.

126. Korn AP, Learman LA. Operations for stress urinary incontinence in the united states, 1988-1992. Urology. 1996 Oct;48(4):609-12.

127. Waetjen LE, Subak LL, Shen H, Lin F, Wang TH, Vittinghoff E, et al. Stress urinary incontinence surgery in the united states. Obstetrics & Gynecology. 2003;101(4):671-6.

128. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. Surgery. 1994 Jun;115(6):707-12.

129. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? Surgery. 1995 Sep;118(3):459-67.

130. Folstein, M.F., Folstein, S.E., McHugh,P.R. “Mini-mental state:” a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189-98.

131. Ward K, Hilton P, United Kingdom and Ireland Tension-free Vaginal Tape Trial Group. Prospective multicentre randomised trial of tension-free vaginal tape and colposuspension as primary treatment for stress incontinence. BMJ. 2002 Jul 13;325(7355):67.

132. Hilton P. Trials of surgery for stress incontinence?thoughts on the ?humpty dumpty principle? BJOG: An International Journal of Obstetrics & Gynaecology. 2002;109(10):1081-8.

133. Maddern GJ, Middleton PF, Grant AM. Urinary stress incontinence. BMJ. 2002 Oct 12;325(7368):789-90.

134. Tennstedt S, et al. Design of the SISTEr (stress incontinence surgical treatment efficacy trial) study: A randomized surgical trial comparing the burch colposuspenion and the autologous rectal fascial sling. Am J Obset Gynecol. In press.

135. McLeod RS. Issues in surgical randomized controlled trials. World J Surg. 1999 12;23(12):1210-4.

136. Finkelstein E, Corso P. Cost-of-illness analyses for policy making: A cautionary tale of use and misuse. Expert Rev Pharmacoecon Outcomes Res. 2003 Aug;3(4):367-9.

137. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. panel on cost-effectiveness in health and medicine. JAMA. 1996 Oct 9;276(14):1172-7.

138. Brown JS, McNaughton KS, Wyman JF, Burgio KL, Harkaway R, Bergner D, et al. Measurement characteristics of a voiding diary for use by men and women with overactive bladder. Urology. 2003 Apr;61(4):802-9.

139. Chang RW, Pellisier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. JAMA. 1996 Mar 20;275(11):858-65.

140. Subak LL, Caughey AB. Measuring cost-effectiveness of surgical procedures. Clinical Obstetrics & Gynecology. 2000;43(3):551-60.

141. Fusco F, Groutz A, Blaivas JG, Chaikin DC, Weiss JP. Videourodynamic studies in men with lower urinary tract symptoms: A comparison of community based versus referral urological practices. Journal of Urology. 2001;166(3):910-3.

142. Griffiths CJ, Harding C, Blake C, McIntosh S, Drinnan MJ, Robson WA, et al. A nomogram to classify men with lower urinary tract symptoms using urine flow and noninvasive measurement of bladder pressure. J Urol. 2005 Oct;174(4 Pt 1):1323,6; discussion 1326; author reply 1326.

143. Grady D, Brown JS, Vittinghoff E, Applegate W, Varner E, Snyder T, et al. Postmenopausal hormones and incontinence: The heart and Estrogen/Progestin replacement study. Obstet Gynecol. 2001 Jan;97(1):116-20.

144. Jackson S, Shepherd A, Brookes S, Abrams P. The effect of oestrogen supplementation on post-menopausal urinary stress incontinence: A double-blind placebo-controlled trial. Br J Obstet Gynaecol. 1999 Jul;106(7):711-8.

145. Grodstein F, Lifford K, Resnick NM, Curhan GC. Postmenopausal hormone therapy and risk of developing urinary incontinence. Obstet Gynecol. 2004 Feb;103(2):254-60.

146. Fultz NH, Herzog AR, Raghunathan TE, Wallace RB, Diokno AC. Prevalence and severity of urinary incontinence in older african american and caucasian women. J Gerontol A Biol Sci Med Sci. 1999 Jun;54(6):M299-303.

147. Wetle T, Scherr P, Branch LG, Resnick NM, Harris T, Evans D, et al. Difficulty with holding urine among older persons in a geographically defined community: Prevalence and correlates. J Am Geriatr Soc. 1995 Apr;43(4):349-55.

148. Thom D, Brown J. Reproductive and hormonal risk factors for urinaty incontinence: Review of the clinical and epidemiologic literature. Journal of the American Geriatrics Society. 1998;46:1411-7.

149. Brown JS, Seeley DG, Fong J, Black DM, Ensrud KE, Grady D. Urinary incontinence in older women: Who is at risk? study of osteoporotic fractures research group. Obstetrics & Gynecology. 1996;87(5 Pt 1):715-21.

150. Brown JS, Grady D, Ouslander JG, Herzog AR, Varner RE, Posner SF. Prevalence of urinary incontinence and associated risk factors in postmenopausal women. heart & Estrogen/Progestin replacement study (HERS) research group. Obstetrics & Gynecology. 1999;94(1):66-70.

151. Diokno AC, Brock BM, Herzog AR, Bromberg J. Medical correlates of urinary incontinence in the elderly. Urology. 1990 Aug;36(2):129-38.

152. Brown JS, Sawaya G, Thom DH, Grady D. Hysterectomy and urinary incontinence: A systematic review.[comment]. Lancet. 2000;356(9229):535-9.

153. Marshall HJ, Beevers DG. Alpha-adrenoceptor blocking drugs and female urinary incontinence: Prevalence and reversibility. Br J Clin Pharmacol. 1996 Oct;42(4):507-9.

154. Fantl JA, Wyman JF, Wilson M, Elswick RK, Bump RC, Wein AJ. Diuretics and urinary incontinence in community-dwelling women. Neurourol Urodyn. 1990;9(1):25-34.

155. Menefee SA, Chesson R, Wall LL. Stress urinary incontinence due to prescription medications: Alpha-blockers and angiotensin converting enzyme inhibitors. Obstet Gynecol. 1998 May;91(5 Pt 2):853-4.

156. Katz S. The index of ADL: A standard measurement of biological and psychological function. JAMA. 1963;185:914-9.

157. Mahoney fi, barthel dw. Functional evaluation: The barthel index. Md State Med J. 1965 Feb;14:61-5.

158. Ouslander J, Johnson T, Nasr S, Schnelle J, Miller M. Atrial natriuretic peptide levels in geriatric patients with nocturia and nursing home residents with nighttime incontinence. Journal of the American Geriatrics Society. 1999;47(12):1439-44.

159. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: The confusion assessment method. A new method for detection of delirium. Ann Intern Med. 1990 Dec 15;113(12):941-8.

160. Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology. 1974 Nov;24(11):1019-25.

161. Wechsler D. Wechsler adult intelligence scale - revised. New York: Psychological Corp; 1988.

162. Homma Y, Ando T, Yoshida M, Kageyama S, Takei M, Kimoto K, et al. Voiding and incontinence frequencies: Variability of diary data and required diary length. Neurourol Urodyn. 2002;21(3):204-9.

163. Groutz A, Blaivas JG, Chaikin DC, Resnick NM, Engleman K, Anzalone D, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: A multicenter study of micturition diary and pad tests. Journal of Urology. 2000;164(3 Pt 1):698-701.

164. Locher JL, Goode PS, Roth DL, Worrell RL, Burgio KL. Reliability assessment of the bladder diary for urinary incontinence in older women. Journals of Gerontology Series A-Biological Sciences & Medical Sciences. 2001;56(1):M32-5.

165. Ouslander JG, Palmer MH, Rovner BW, German PS. Urinary incontinence in nursing homes: Incidence, remission and associated factors. J Am Geriatr Soc. 1993 Oct;41(10):1083-9.

166. Colling J, et al. Continence program for care dependent elderly (final report). 1995. Report No.: NIH-NCNR-NR01554.

167. Siltberg H, Victor A, Larsson G. Pad weighing tests: The best way to quantify urine loss in patients with incontinence. Acta Obstet Gynecol Scand Suppl. 1997;166:28-32.

168. Griffiths DJ, McCracken PN, Harrison GM. Incontinence in the elderly: Objective demonstration and quantitative assessment. Br J Urol. 1991 May;67(5):467-71.

169. Farhat W, Bagli DJ, Capolicchio G, O'Reilly S, Merguerian PA, Khoury A, et al. The dysfunctional voiding scoring system: Quantitative standardization of dysfunctional voiding symptoms in children. J Urol. 2000 Sep;164(3 Pt 2):1011-5.

170. Sureshkumar P, Craig JC, Roy LP, Knight JF. A reproducible pediatric daytime urinary incontinence questionnaire. J Urol. 2001 Feb;165(2):569-73.

171. Campbell MF. Campbell's urology. 8th ed. Philadelphia, PA: WB Saunders Company; 2004.