The year of 2012 was a very exciting year for the Society for Fetal Urology (SFU). It was highlighted by two outstanding national meetings, a historic management agreement reached between the SFU and the American Urological Association (AUA), and the execution of critical memorandum of understanding (MOU) between the SFU and the Society for Pediatric Urology (SPU) as regards the role of the SFU in the SPU/AUA Annual Meeting.

The 48th Bi-Annual Meeting of the SFU was held in conjunction with the AUA Annual Meeting in Atlanta, GA. The full day program was coordinated by Dr. Armando Lorenzo (The Hospital for Sick Children, Toronto, ON). The meeting was centered on “study design, data analysis, and reporting in prenatal and perinatal urology.” The meeting was organized with ten mini-symposia by experts in the field including: “Randomized Controlled Trials” by Dr. Edwin Smith (Georgia Urology, Atlanta, GA), “Non-Experimental/Observational Studies” by Dr. Nicol Bush (Children’s Medical Center, Dallas, TX), “Population-Based Research” by Dr. Katherine Herbst (Connecticut Children’s Medical Center, Hartford, CT), “Systematic Reviews and Meta-analysis” by Dr. Luis Braga (McMaster Children’s Hospital, Hamilton, ON), “Clinical vs. Statistical Significance” by Dr. Hillary Copp (UCSF Benioff Children’s Hospital, San Francisco, CA), “Guidelines, Consensus Statements and Recommendations” by Dr. Bob Nguyen (Boston Children’s Hospital, Boston, MA), “Ethical Issues in Research” by Kourosh Afshar

As you will read in John Kryger’s summary, the Society for Fetal Urology has had a busy year academically and politically. Dialogues in Pediatric Urology is pleased to continue to showcase their presentations in this annual Society for Fetal Urology Special Edition.

For those of us who are indeed hung up on the p-value, read on! This Edition of Dialogues features four contributions from Armando Lorenzo’s excellent morning program from the Spring SFU meeting. Since this is a special edition, we were also able to showcase a few of the Spring case presentations. Part II will feature the abstracts from the Fall meeting.

Thank you to each of the authors for sharing their expertise so thoroughly and clearly—the contributions are truly accessible to the less indoctrinated scholars who want to do honest, worthwhile clinical research. There is an enormous amount to learn and practice. I had no idea that there were so many forms of bias that affect every aspect of our work! It is also very clear how much work goes into creation and execution of a solid prospective RCT—hats off to those who take on the challenge.

Now that many Pediatric Urology fellowship programs are offering formal additional coursework, training or degrees in clinical investigation, our field certainly has a bright future for high quality clinical research.
From the SFU President

(British Columbia Children’s Hospital, Vancouver, BC), “Publishing in High-Impact Journals” by Dr. Anthony Caldamone (Hasbro Children’s Hospital, Providence, RI), “Manuscript Review” by Dr. Paul Mergerian (Seattle Children’s Hospital, Seattle, WA). Finally, Dr. Anthony Herndon (University of Virginia Pediatric Urology, Charlottesville, VA) outlined the “SFU research goals” and reviewed the current and future projects and registries. The meeting finished with a 2-hour session of case presentations on topics of fetal urology.

The 49th Bi-Annual Meeting was held in conjunction with the American Academy of Pediatrics (AAP) National Conference and Exhibition in New Orleans, LA. Dr. Eric Jones (Houston Pediatric Urology, Houston) coordinated the program. It was highlighted by a lecture on “Ethical Dilemmas in Fetal Urology” (Dr. David Diamond, Boston Children’s Hospital, Boston, MA), “Fetal Urologic Imaging” (Dr. Amy Mehollin-Ray, Texas Children’s Hospital, Houston, TX), “Open Fetal Surgery for Myelomeningocoele Closure” (Dr. Michael Babington, Director of Prenatal Diagnosis and Fetal Imaging, Texas Children’s Fetal Center, Houston, TX), “Marks Fetal Surgery for Congenital Malformations” (Dr. Darryl Cass, Director of Texas Children’s Fetal Center, Houston, TX), “Fetal Therapy for Obstructive Uropathy - A 30-year Retrospective” (Dr. Andrew Freedman, Cedars Sinai Medical Center, Los Angeles, CA), “Importance of Multidisciplinary Team in Optimizing Fetal Care” (Dr. Pramod Reddy, Cincinnati Children’s Hospital, Cincinnati, OH). These were followed by a 3-hour session of case presentations and studies pertaining to fetal urology.

We were proud to ratify a historic management agreement with the AUA. This will bring us into a secure relationship with the AUA in a fashion consistent with other AUA-affiliated societies. The agreement will include official headquarters in the offices of the AUA along with dedicated management staff. It will strengthen the financial services, meetings management, marketing, industry relations, and membership services to the SFU.

In addition, a great deal of time and effort was dedicated to drafting an MOU that would ensure the SFU is able to continue bi-annual meetings in conjunction with the SPU in the future. This was an especially critical initiative, since pediatric urology has withdrawn its annual fall meeting from the management of the AAP. Thus, it is a moment of great change and restructuring in the world of pediatric urology and an opportunity for improvement. We have worked hard to maintain dedicated meeting time for the SFU as part of the SPU/AUA Annual Meeting in the spring and the new Pediatric Urology Fall Congress. The agreement protects the autonomy of the SFU as regards meeting planning and ensures representation on the planning committees.

The successes of this year would not have been realized without the tremendous effort of Dr. John Weiner (Duke Children’s Hospital, Durham, NC), who was SFU President, and Dr. Jeffrey B Campbell (Children’s Hospital Colorado, Denver, CO), who was SFU Secretary/Treasurer. They worked closely with Dr. Gopal Badlani (Wake Forest School of Medicine, Winston-Salem, NC), AUA Secretary and Drew Shifflet, AUA Director of Committee and Society Affairs to establish the management agreement with the AUA, and Dr. David Diamond (Boston Children’s Hospital, Boston, MA), SPU Past President, Dr. Larry Baskin (UCSF Benioff Children’s Hospital, San Francisco, CA), SPU President, and Dr. Ross Decter (Penn State Hershey Surgical Specialties, Hershey, PA), SPU Secretary to establish the MOU between the SFU and the SPU. Furthermore, we are indebted to the insight and efforts of Dr. Anthony Herndon and Dr. Chris Cooper (University of Iowa Pediatric Urology, Iowa City, IA) in forging these relationships.

I look forward to ongoing success in 2013. I want to congratulate and thank Dr. John Wiener and Dr. Jeffrey B Campbell who completed their terms in office of the SFU. The year of 2012 will be a significant year in the history of the SFU due to their exceptional efforts. I welcome Dr. Armando Lorenzo as the new SFU Secretary/Treasurer and Dr. Jeffrey B Campbell as the SFU President-Elect to continue our trajectory of success along with myself as the current President of the SFU. Plans are underway for the 50th Bi-Annual Meeting, which will be a half-day session in conjunction with the SPU/AUA Annual Meeting in San Diego, CA. I hope our 50th will be momentous!

We are grateful to the Dialogues in Pediatric Urology for the opportunity to feature the proceedings of our society. Additional information can be found on the SFU website: www.sfu-urology.org.
Randomized Controlled Trials: The Frequently Recommended but Rarely Achieved Gold Standard

Arun Srinivasan, M.D., Children’s Hospital of Philadelphia
Edwin Smith, M.D., Children’s Hospital of Atlanta

What makes randomized clinical trials (RCT) the “gold standard” in study methodology and how frequently are they currently employed?

For centuries physicians have relied on careful, practical observation to make deductions about the effectiveness of treatments. Historically, the introduction of interventions has been by development of a procedure by a surgeon or group of surgeons, followed by observation and report of results in a case series or retrospective or prospective cohort study. Observation studies will always be a source of evidence when: 1. treatment effect is large; 2. only one acceptable treatment is available; 3. uncommon diseases are being studied’ and 4. when funding is difficult. Yet, in modern medicine the treatment effect of a new technology is often small and causality of outcome by observation alone is less certain. When the treatment effect is likely smaller, extraneous factors must be controlled to be certain that the observed difference is indeed due to the newer treatment or treatment modification. RCTs involve the random allocation of participants or subjects to two, sometimes more, treatment groups that are believed to be equal in both known and unknown attributes. Except for the interventions being compared the patient groups are managed and followed in an identical manner, prospectively, with the outcome of interest to be measured defined a priori. However, a critical point is that the patient group being randomized is only a representative sample of the population of interest.

RCTs find their context within the realm of evidence based medicine (EBM), which is defined as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”. It is expected that there will be integration of clinical expertise and judgment along with patient and societal values with this best available evidence. Less encumbered with bias by design, systematic reviews of RCTs, followed by individual RCTs, should offer the most valid and reliable results. Therefore, they occupy the pinnacle of strength within evidence based medicine. While studies of an observational nature can also provide important evidence in guiding clinical care, they are assigned a lower level of evidence. Therefore, they occupy the lower rung within evidence based medicine. While studies of observational nature can also provide important evidence in guiding clinical care, they are assigned a lower level of evidence.

What is the proper ethical foundation for designing, recruiting, enrolling and maintaining a patient in a randomized clinical trial?

The primary duty of any physician is to act in the best interest of their patient - selecting with deliberation the individualized treatment that he or she believes to be the one most likely to improve the welfare of the patient. How then can we address the apparent conflict of interest when patients are enrolled in a study and exposed to a therapy that is selected at random with an uncertain relative benefit, especially when risk is involved? When a physician becomes an investigator there is a potential for conflict of interest that follows the desire for recruitment and maintenance of enrollment of patients until the study is completed. This is a concern that is both acknowledged and addressed through the principle of clinical equipoise. It is an ethical prerequisite for the conduct of any randomized clinical trial.

Equipoise is consistent with the expectation that clinical research trials begin with an honest null hypothesis. Medical ethicist Benjamin Friedman’ explains that clinical equipoise exists when “there is a genuine uncertainty within the expert medical community – not necessarily on the part of the individual investigator – about the preferred treatment”. There is no need for a trial if one treatment is clearly superior. When there are multiple arms to a trial, equipoise should exist between each arm. Finally, placebo controlled trials are legitimately employed under the special circumstance of there being no known effective treatment for the condition under study.

Friedman goes on to explain that “the trial must be designed in such a way as to make it reasonable to expect that, if it is successfully completed, clinical equipoise will be disturbed. In other words, the results of a successful trial should be convincing enough to resolve the dispute among clinicians.” During the course of a study, a clinician or safety monitoring committee may accumulate sufficient evidence so that any expert physician in the field would reach a conclusion that one treatment is better. Alternatively, the complication rate of a treatment may prove unacceptable. When this occurs clinical equipoise has been disturbed and trial is terminated even if well short of the expected schedule. As an example, patient accrual for the MOMS study design called for the enrollment of 200 women, 100 per arm. Yet, it was halted after interim analysis of the data revealed that the predetermined statistical criteria for stopping the study had been achieved. A clinical benefit to in utero closure of the spine had been demonstrated after 183 patients had been randomized. Clinical equipoise therefore governs all aspects of RCT design.

What is the “Achilles Heel” of a randomized clinical trial?

There are two processes invariably present that threaten the clinician’s effort to design and conduct a RCT: bias and chance. In clinical trials, bias is usually unintentional and often unrecognized and often it is systematically introduced through a difference in the way study groups are assembled or in the way they are measured. Bias leads to either an underestimation or overestimation of the effects of intervention. Chance is inherent in all observations; it is the result of ran- (continued on next page)
Randomized Controlled Trials

Randomization means that chance alone determines treatment assignment. But bias in various forms may be introduced and confound results. It is important to understand the various types of bias and to design the study to offer protection against the introduction of bias.

Selection Bias

A selection bias would exist if a comparative study of TIP vs onlay hypospadias repair were conducted and the TIP patients all had a wide deeply grooved urethra and the onlay patients did not. A robust randomization process ensures that the allocation of “difficult” or of “easy” subjects to one treatment over the other is purely by chance. This offers the potential to isolate and to quantify the impact of an intervention.

Fundamental to randomization is a uniformly exercised allocation sequence that participants, study personnel, investigators will not be able to predict. In the past opaque sealed envelopes were a common method but have disadvantages. Other problematic methods of assignment include date of birth, clinic day or medical record number, which are systematic occurrences rather than random events and are thereby vulnerable to predictability. Distance randomization is preferred. Both the MOMS trial and the RIVUR study employed a web-based data management system. Once the sequence is defined it must be diligently followed in order to eliminate the possibility of determining ahead of time the group to which a patient will be assigned. Blocked randomization may be used to reduce the risk that different numbers of people will be randomly assigned to the treatment (T) or control (C) groups. Patients are randomized by blocks. For example, with a fixed block size of 4, then patients can be allocated in any of the orders: TTCC, TCTC, CTCT, TCTC, CTTC, or CCTT. The order is chosen randomly at the beginning of the block.

It may be important to ensure that the treatment and control groups are balanced on important prognostic factors that can influence the study outcome (e.g., gender, ethnicity, age, socioeconomic status or, in a vesicoureteral reflux trial, the grade of reflux). This is where patient selection criteria play an important role. Selection criteria for a study should do the following: a. keep the study sample a representative sample of a wider population and b. avoid inclusion or exclusion criteria that would shift the outcomes with regards to the intervention studied. Before doing the trial, the investigator decides which strata are important and how many stratification variables can be considered given the proposed sample size. A separate simple or blocked randomization schedule is developed for each stratum. An interim analysis should be part of the strategy to ensure that the study groups are comparable.

Ascertainment Bias

This occurs when the results and conclusions of a trial are systematically distorted by knowledge of which intervention each participant will be or is receiving. So while allocation concealment protects the randomization sequence before allocation, ‘blinding’ and completeness-of-follow-up makes sure they remain comparable except for intervention-of-interest during the course of the study. Ideally, all individuals participating in a trial should be blinded but this may not be feasible. In the past the following designations were followed. When the subjects receiving intervention are blinded, the trial is single blinded. When individuals administering interventions are blinded, the trial is double blinded. When the individuals in charge of assessing and recording outcomes are blinded, it is triple blinded. However the 2010 CONSORT statement suggests that blinding terminology be simplified to identifying the groups of individuals (participants, healthcare providers, data collectors, outcome adjudicators, or data analysts) that were blinded in order to prevent bias in a trial through knowledge of the treatment assignments.

Blinding would be desirable for surgery, which as an intervention often has a significant placebo effect. For obvious reasons blinding is challenging to apply to surgical trials. Sham surgery violates ethical barriers, the surgeon cannot be blinded to the treatment, and the incision is apparent to the patient.

Observational Bias

Observational bias may happen when an outcome may be estimated differently by different people making the observation. Examples include estimation of Gleason score in prostate cancer studies, identification of renal scars in reflux studies, estimation of pain in outcome studies after surgery. Common elements to all these factors are that they are observations that can be variably made by different people in an honest fashion and hence the difficulty in prevention. The means to decrease or avoid observational errors are as follows: a. use of standardized tools (for instance, use of validated questionnaires in evaluating interstitial cystitis patients or visual analog pain score for assessing pain are tools that remove the observer’s bias and thus decrease this type of error) b. defining the observation before the start of the study and ensuring that all observers adhere to these definitions by using interim analysis and random analysis. It is also very important to specifically define these measures when describing the study and reporting outcomes as it goes to validation of the study and its conclusions.

Other Bias

The sources of bias are nearly infinite. Just a few examples: in
Randomized Controlled Trails

choosing a research question researchers may be influenced by cost, funding availability, or hold a vested interest in outcome. Population bias occurs when the study group is overly restricted to exclude certain members of the population. Intervention bias may be introduced when the intervention is new, the technique is evolving, or the surgeon has a learning curve. Withdrawal bias occurs when withdrawals from the study are not properly accounted for. Various consumer biases exist, such as famous author bias, famous institution bias, or prestigious journal bias. Careless reading bias occurs when we only read the abstract and conclusions. Measurement bias is introduced when outcomes are chosen that are easy to measure rather than clinically relevant. More often than not, bias is completely unintentional but nevertheless destructive to strength of a study’s conclusions. To summarize we recognize bias as “any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth.” 

Why do we need to apply inferential statistics to the observed results of a RCT?

Chance happens. It is inherent in all observations. It can be minimized, but never avoided. It occurs as the result of the random variation of a biologic phenomena or it may occur as investigators attempt measurements of the phenomena. Because research studies are conducted with a patient sample, we need inferential statistics to bridge the gap between observed sample results and hypothesized universe states. In contrast to retrospective observational studies, it is axiomatic to RCTs that the statistical analysis is preplanned. We must be able to construct the study to limit the effects of chance and protect internal validity to safeguard the potential for generalizability of the results.

Based on the defined primary outcome, the principal conclusions of a trial are usually expressed in dichotomous terms—either the treatment was effective or ineffective. But there are two instances when erroneous conclusions can be drawn and they are as follows:

a. When the error is an alpha or Type I error—which is analogous to a false positive—we conclude effectiveness when the treatment is ineffective. Therefore, a Type I error falsely rejects the null hypothesis. Diagnostics and pharmaceutical companies, device manufacturers, innovative surgeons and trial sponsors want to avoid a Type I error, which would lead them to focus resources on development of a drug, medical device or procedure that will ultimately prove to be a failure.

b. Beta or Type II error is analogous to a false negative—concluding the treatment is ineffective when it is in fact effective. A negative trial will likely discourage interest and funding development of a new diagnostic tool or treatment. When a Type 2 error occurs the investigator is falsely led away from productive avenue of research. This is not only costly to the sponsor and investigator but it is costly to society, which loses out on finding an effective diagnostic study or treatment.

Inferential statistics offer the means to calculate the likelihood of arriving at an incorrect conclusion. The chance of an alpha error is expressed as the familiar p value. It is a quantitative statement of the probability that the observed differences in the study could have happened by chance alone assuming that there is in fact no difference between the groups. Beta is the probability of declaring no difference when a difference exists. Conversely, Power equals 1 – beta and is the probability of detecting a difference when one exists. It is analogous to sensitivity. In clinical trials people will tend to focus on whether p<0.05. They want to know that there is a less than one chance in 20 of the study producing a false positive result. Ironically, the likelihood of a beta error and power are often lost in the shuffle. Yet, power is the probability that the study results will achieve a p<0.05 and therefore satisfy everyone. Having sufficient power ensures success. Yet, power comes at a cost: sample size. So although unlimited power would be desirable, the need for sufficient power must be balanced against the cost of sample size.

What are the study and statistical essentials of a good RCT?

Primary Outcome

While a surgical RCT may have a number of outcomes (surgical, clinical, patient reported QOL, economic) selected to address the corresponding objectives of the study, there must only be one primary outcome. The primary outcome should be capable of providing the most clinically relevant evidence related to the aim of the study and should reflect the accepted norms and standards in the relevant field of research—what is important and how will it be measured as a consensus. It must be decided at the outset of the study and serve as the basis to calculate power.

Selection Criteria

Inclusion and exclusion criteria form the back bone of defining the study population. As noted earlier, selection criteria impact the extension of the study conclusions to a wider population and for this reason should be a representative of the population for whom the conclusions will be applied. At the same time it should not introduce variables by virtue of patient selection that will bias the results one way or another.

Sample Size Calculation

Otherwise called the power analysis, this is the estimated sample size to demonstrate an effect that is within a predetermined range. The logical stepwise process should be clearly explained as it is important to avoid a Type II error.

Randomization Process

The process should be clearly explained and rigorously followed to ensure validity.

Pre study comparison of groups: A RCT report will include a table comparing the baseline variables between the two groups. At the least, the groups should be not statistically different and hence comparable. This comparison also serves as a means to demonstrate that the randomization process has worked well.

As a caveat, Altman was one of the first to point out that a variable that is not statistically different between the groups could still influence outcomes. Hence two groups cannot be deemed comparable or non-comparable just based on the initial table showing a set of variable between the two groups and their p value. Hence when comparing two groups, Senn recommends identifying variables influencing outcomes before the start of the study, adhering to these in the analysis, and avoiding the addition and speculation of impact of a “surprise variable” discovered during the study.
Randomized Controlled Trials

Numerical Analysis

Addressed in a later question.

Limitations and Generalizability

Investigators must identify the patient population that will be affected by the results of the study. Obviously, the more restrictive the inclusion/exclusion criteria employed the more limited the population for whom the results may be applied. An appraisal of the potential limitations of a study should be recognized before its initiation and revised as the study progresses. Specifically, factors that might compromise the precision of measurement of the primary outcome along with identification of real or potential biases should be identified at inception, monitored during the study and discussed in the study manuscript.

What are the important steps in data analysis when evaluating a RCT?

Bias is the single biggest confounder in a study since it can inexplicably influence results in spite of the author’s best intentions. Randomization is considered superior since it supposedly limits selection bias. But in the process of randomization other biases might be introduced and one has to be wary of those issues when analyzing a RCT.

We suggest a simple checklist to assess a study’s methodology as follows:

1. Ensure:
   a. a well defined patient population with clearly defined inclusion/exclusion criteria
   b. definition of clear primary and secondary end points that are clinically relevant and make common sense
   c. selection of an appropriate randomization technique

2. Establish a protocol of analysis – “intention to treat” or “per protocol”.

   “Intention to treat” principle is to analyze subjects based on their initial intention of treatment assignment regardless of the change of treatment or change of mind at a later time. This is done to preserve the sanctity of the randomization process from the effects of subject cross over or drop out which might introduce ascertainment and observational bias. On the other hand “per protocol” keeps patients in groups based on what really happened rather what was intended. Statisticians have long argued the merits and demerits of these two methodologies which are beyond the scope of this discussion. Intention to treat is considered the standard in randomized controlled trials. It mostly underestimates the efficacy of treatment since it considers protocol violators as active participants and thus dilutes success rate (which is better than inflating success rate). Authors should establish a protocol and adhere to the protocol through to the completion of the study.

3. Make sure that there are not a large percentage of subjects who are drop outs, non adherence to protocols and lost to follow up.

4. Sample size – while there is a minimum for meaningful results (based on an appropriate power calculation), there is no maximum (at least not in the world of medicine). Larger sample size mostly implies more reliable results. In other words, in sample size - bigger is better.

5. Groups are compared using appropriate statistical techniques. Here are some commonly used (but over simplified here) statistical techniques that most authors agree are appropriate for a RCT:

   a. For binary outcomes – (cured or not cured) - Logistic regression
   b. For continuous outcome – (improvement or not) - Linear regression
   c. Time to event analysis – Kaplan-Meier estimate and Cox proportional hazards analysis

6. A randomization test to provide a “model free” estimate of the statistical significance is optimal. This establishes that the randomization process has worked and adds validity to the results.

7. Follow up period particularly for time to event analysis should be appropriate to obtain relevant data. Censoring occurs when a patient is removed from the survival curve at the end of their followup. Hence outcomes are only partially known, and the sample size beyond the removal of the censored patient is reduced. For instance, when a bladder cancer study looks at survival data but provides follow up for patients for only two years the reliability of the study is compromised. The more patients that are censored, the more unreliable the results.

8. Sub-group analysis in a randomized controlled trial should be generally discouraged, as the primary study was not designed to study the sub-group in question. The study often will lack the power to address the question in a statistically sound manner.

But most importantly please remember, statistics can trump common knowledge and opinions but cannot trump common sense and clinical knowledge.

Is there a standard to which randomized controlled trials are held? Who designed these standards?

CONSORT (CONsolidated Standards of Reporting Trials) was developed by a group of clinical trialists, statisticians, epidemiologists and biomedical editors and was first published 1996. The CONSORT statement was offered in response to the poor quality of many published randomized controlled trials and the absence of minimal reporting standards. It sets out the standards of structuring, analyzing and reporting a randomized controlled trial. The statement consists of recommended checklists and flow diagrams to report a randomized controlled trial and is available online (http://www.consort-statement.org/).

Since the CONSORT statements release it has been updated twice – 2001 and 2010. The standards have been adapted by multiple institutions and journals, and preliminary data support adherence to CONSORT statement improves reporting quality. The statement is seen as guidelines rather than strict protocols that set the minimal standards of reporting but can be improved by publishing authors as they deem fit. Preliminary studies show improved reporting quality following publishing of these guidelines.

What are the challenges particular to conducting surgical RCT?

Several obvious barriers to the performance of surgical RCTs exist which help to explain the low representation of this methodology in the surgical literature. Double blinding is impossible as the surgeon must know the treatment, and a placebo group for the patient would require sham surgery that exposes the patient to significant risk without benefit. Additional challenges include the irreversibility of surgical treatment, appropriate timing for assessment of a newly introduced procedure, potential subjects’ lack of comprehension, data collection and expense.
Randomized Controlled Trails

Timing of Study

Timing of conducting a surgical RCT often presents a dilemma and it can be too early or too late. All surgeons agree that a learning curve exists for procedures. With an RCT introduced too early there is a risk of recognizing the need for procedure modifications. Yet, if modifications are introduced during the trial, the principle of uniformity in intervention is violated. However, without these modifications the procedure under investigation becomes obsolete before the trial is completed. Too late and the procedure may already be “accepted” by the community without trial validation.30

Patient Participation, Choice and Understanding

When patients learn about plans to investigate a novel medicine or surgical procedure, especially if alternative treatments seem undesirable, it is not uncommon for investigators to receive inquiries as follows: “I want (or I want my child) to participate but I want (or want him/her) to get the surgery.” The vast majority of prospective patients will not understand the structure of RCTs and will not understand what it means to participate. When initially interviewed, patients recruited to urologic RCTs have answered with regard to what trial meant that “only the newer procedure was being tested.” Half of them could not accept that physicians could be uncertain about which treatment is best and still believed that the physician would choose which intervention best suited their specific medical condition. Some rationalized that randomization moves responsibility of choice from physician and provides means for hospitals to ration expensive care. It is important to recognize that across the board in RCTs, an average of 60% of eligible patients will agree to participate and up to 50% will drop out.29 Patient choice can also confound structuring a surgical RCT. Randomization might not be possible due to patient perception of presumed benefits. If most patients in a study really prefer to be in one particular arm of the study, it increases the cross over risk, increases risk patient dissatisfaction leading to drop out and other such problems.30

Funding

A randomized surgical trial can be expensive and it might be challenging to obtain funding since they are not typically incentivized by the private industry. Historically, with regard to National Institute of Health funding, surgical RCTs have been less likely to receive funding, and awards have been of lesser value.29-31 Recently, the NIH has shown more interest in outcomes research, including surgical outcomes.

Data Collection

Outcomes studied after a surgical intervention can be difficult to define. Before the study starts, all outcomes should be specifically described in a fashion that embraces common sense and medical knowledge.28, 30 For example, when presenting data on improvement in continence, the investigator’s opinion about acceptable results may not be agreed by all. Hence, one should use a validated method to assess and estimate continence both before and after surgery with clear definitions rather than impressions. Rather than indicating that continence improved, more objective information of how many pads patient used before and after surgery should be reported. Likewise pain assessment, aesthetic outcomes and quality of life are vulnerable to subjectivity and imprecise measurement, making comparison of outcomes from similar studies difficult to compare.

What are the limitations of RCTs and can they be addressed by other trial designs?

Randomized controlled trials are rightfully considered the best statistical tool to address a research question due to their ability to assign causality through the absence of selection bias and random distribution of confounding factors across treatment groups. But by no means should they be considered the only effective tool. Concato et al compared outcomes from RCTs and observational studies looking at five clinical topics using a meta-analysis.31 They found similar outcomes and conclusions in both the RCT and observational studies. They concluded that a well designed observational study should not be discounted solely on the perceived value of its methodology.

Although observational studies are more vulnerable to flaws such as selection bias, confounding variables and inadequate sample size, under certain circumstances they may be the most appropriate study method. RCTs are not always feasible due to cost, time involved, and the rarity of the disease process or the complicated logistics necessary for the study.32 In addition, an RCT may not be possible when there are ethical barriers to randomization or the observation period to measurement of outcomes is lengthy. Finally, inclusion and exclusion criteria may create a sample population that is remote from the diversity found in the general population. In these situations, observational studies may prove very helpful to provide valuable evidence that can guide treatment. Any trial data and conclusions are only as good as the design of the study. Hence a poor study design would negate the advantage of RCT.33

References

Randomized Controlled Trails

(continued from previous page)


Clinical Versus Statistical Significance: Let’s Get Past the P-Value

Hillary Copp, MD
UCSF Benioff Children’s Hospital, San Francisco, CA

We use statistics to determine whether an observed outcome is the result of a true relationship between two variables or the result of a chance occurrence. There are two basic statistical methods to assess the role that chance plays in a relationship: p values and confidence intervals. What does each of these methods tell us? Is one method preferred over the other?

In this overview we will cover: 1) The p value, 2) The confidence interval, 3) The association between the p value and the confidence interval, and 4) Choosing between the p value and the confidence interval.

What is a P Value?

The p value is the probability that the observed difference between compared groups is due to chance alone. By convention, when p < 0.05 we say it is very unlikely that the difference observed is due to chance. We apply the p value through hypothesis testing. In hypothesis testing we formulate a null hypothesis into a negative statement (i.e. the two treatments are not different). We then perform the statistical test to determine the probability (p value) that the observed difference is due to chance. If the probability is less than 5% (p<0.05) we are willing to reject our null hypothesis and say the difference is real.

As an example let’s look at a hypothetical weight reduction medication called Weight Loss Pill # 1. A study was done comparing people who took Weight Loss Pill # 1 with those who did not. The study showed no significant difference in weight loss for people who took Weight Loss Pill # 1 (p=0.06). With p=0.06 we are not willing to say that the difference in weight reduction between the two groups is real because the likelihood that the results are due to chance is greater than 5%. Therefore, this study showed no statistically significant difference in weight loss for people on Weight Loss Pill # 1 versus controls.

Confidence Interval (CI)

The confidence interval represents the range of effects that might be expected based on the data. Overtime it has become convention to report the 95% confidence interval, although other values could be used such as 99%. A confidence interval of 95% indicates that 95 out of 100 times the range of values contains the true value. Another way of saying this is that we are 95% confident that the true difference lies within the interval. With the 95% confidence interval we are given the measured effect (also known as the point estimate) which is the actual measured difference between the two groups. The 95% confidence interval then tells us how confident we are in the measured effect.

As an example let’s look at another weight reduction medication called Weight Loss Pill # 2. The point estimate with 95% confidence interval for those who take Weight Loss Pill # 2 is -0.5 (95% CI -1 to -0.2). This means that the measured difference between those who took Weight Loss Pill # 2 compared with those who did not was a half a pound and that we are 95% confident that the amount of weight reduction using this medication could be as much as one pound to as little as 0.2 pounds.

The Association between the P Value and the Confidence Interval

The 95% confidence interval contains information about the p value. From the confidence interval alone we can tell whether the measured effect between the groups is statistically significant (p<0.05). If the confidence interval does not overlap the area of “no difference” then the result is statistically significant. Consider the example using (continued on next page)
Clinical vs. Statistical Significance (continued from previous page)

the two weight reduction medications below (see Figure 1). For Weight Loss Pill # 1 the 95% confidence interval includes the area of “no difference” and therefore we would expect the p value to be > 0.05. For Weight Loss Pill # 2 the 95% confidence interval does not overlap the area of “no difference” and therefore we would expect the p value to be < 0.05.

The effect measure (means, risk ratios, etc.) used for comparison determines the value for “no difference”. With means and proportions, subtraction is used to determine the difference between groups. Therefore, the value for “no difference” is zero. With relative risks or odds ratios division is used to determine the difference between groups. As such, “no difference” is equal to one. In the example above the studies are comparing differences in mean weight loss and as a result “no difference” is equal to zero (no mean weight loss or gain).

Choosing between the P Value and the Confidence Interval

Significant p values tell us what we observed is unlikely to be due to chance; however, the p value alone can be misleading. For example, although Weight Loss Pill # 2 causes a statistically significant weight reduction between those who take the medication and controls, the average mean weight loss is only one half pound. Is this quantity of weight loss clinically meaningful? This is an example of a difference that is statistically significant, but the difference is so small that it is not clinically relevant. In addition, p<0.05 is completely arbitrary. It is important to remember that probability is a continuum. Is there really a difference between p < 0.05 and p < 0.06- a 5% versus 6% chance that the difference observed was real?

On the other hand, confidence intervals give the magnitude of the effect (the point estimate), the range of effects, and whether or not the observed difference between the compared groups is statistically significant. Thus, the confidence interval is the preferred method for reporting results of statistical analysis. That said reporting the p value along with the confidence interval, while not necessary, is ok. However, there are situations where sole use of the p value is appropriate such as in reporting the p value from a univariate analysis that was used to select covariates for multivariable modeling. Another common instance where solo p value reporting is appropriate is within large tables in manuscripts or PowerPoint presentations where placing the 95% confidence interval in for every cell within the table is cumbersome and can make the table illegible.

Going beyond Statistical Significance

It is important to remember that there are questions that neither p values nor confidence intervals can answer. First, a significant p value does not tell one whether the observed result is clinically significant. This requires interpretation of the results within a clinical framework. For example, although Weight Loss Pill # 2 had a statistically significant weight loss of 0.5 pounds this is not necessarily a clinically relevant weight loss. Second, is the observed result from the analyzed data reflective of the true population? This will depend upon whether a representative sample of the population is examined. As an investigator it is critical to do everything you can to make sure you understand who your study population is. Moreover, when you are attempting to apply the literature to your individual practice, ask yourself if the sample population is generalizable to the patients to whom you deliver care. Finally when reviewing the literature, analyzing your own study results, or interpreting abstracts and data presented at conferences ask yourself the following four questions to place the study results into a meaningful context: What is the magnitude of effect? Are the results statistically significant? Who makes up the sample population? Do the results have clinical relevance? If the answers to these questions are acceptable then the study in question is likely of clinical significance.

References

Systematic Reviews and Meta-Analysis: Garbage In Garbage Out versus Meaningful Data Crunching

Luis H. Braga MD MSc PhD
Associate Professor – Department of Surgery/Urology and Epidemiology and Biostatistics
McMaster University, Hamilton, Ontario, Canada

Terminology

Systematic reviews (SR) are reviews that incorporate strategies to minimize bias by having clearly formulated research questions, pre-defined inclusion and exclusion criteria and explicit methodology to identify, select, and critically appraise studies. These key characteristics of SR set them apart from traditional narrative reviews. The main differences between narrative reviews and SR are summarized in the Table 1.

SR may not necessarily include a meta-analysis, which can be defined as the statistical pooling of data from individual studies to create a single treatment effect estimate (pooled estimates of effects). In a meta-analysis, combining individual studies of small sample size increases statistical power (by increasing the total sample size of the study) and generates more precise effect estimates for that particular inter-

vention under investigation. Usually, meta-analysis is the final step of a SR, and should only be performed when the variation of treatment effects between studies is small (low heterogeneity), allowing their pooling into a single estimate.

Importance

When properly conducted, SR and meta-analyses of randomized controlled trials are ranked high in the hierarchy of evidence and may serve as the foundation for evidence-based practice guidelines, helping clinicians, researchers and policy-makers with a synthesis of the fast-growing literature on the field. Their results can therefore influence clinical decisions, future research agendas and economic evaluations.

(continued on next page)
Nevertheless, like any other research study, particularly one that is observational and retrospective, SR and meta-analyses can be subjected to systematic (bias) and random (chance) errors due to their own limitations. These intrinsic limitations may originate from poor quality of included articles, heterogeneity between studies, and presence of publication and outcome reporting biases. Therefore, the validity and applicability of any SR depends mainly on the quality of the included studies and on how the review was carried out. Despite that, SR are currently the best, least biased, and most rational way to organize, select, evaluate, and summarize the research evidence from the expanding medical literature.

Steps in conducting SR (Table 2)

1. **Research Question**

   Every SR is based on a research question. To formulate a research question, one should consider the following aspects related to the research topic (PICO): Patient, population (P): Which characteristics (clinical condition, age group) define the population? Intervention (I): Which intervention or exposure (type of treatment – medication or surgery, diagnostic test, or educational program) is applied to the patient/population? Comparison intervention (C): Which intervention is the treatment under study being compared to? Outcomes (O): Which outcomes are being studied (complication, mortality, quality of life, cost-effectiveness)?

2. **Inclusion and Exclusion Criteria**

   Inclusion criteria for types of patients and interventions under investigation must be clearly defined before data abstraction to avoid inclusion criteria bias, i.e. modifying the inclusion criteria towards a direction of effect after reviewing the results of relevant studies.

   Controversy exists regarding what type of study design should be included in SR. Observational studies should not be combined to randomized clinical trials in SR whenever enough randomized trials can be found on that particular topic under study. In contrast, if the aim of a SR is to define the current state of knowledge on a particular field, then including the “best studies available” is appropriate even if these best studies are not randomized clinical trials.

3. **Search Strategy**

   The defining characteristic of SR is the comprehensive and systematic search used to identify all studies that form the body of evidence relevant to the PICO format of the research question. The validity of the results of any SR is invariably linked to the search strategy employed to find this evidence; therefore it is essential to perform an exhaustive literature search. From an initial list of citations, relevant articles should be selected based on predefined inclusion and exclusion criteria, regardless of their results or quality. For reproducibility, it is important to report each step of the applied search strategy and its results in detail in the methods section.

   A quick search only including MEDLINE is generally not considered adequate. Studies have shown that only 30% to 80% of all known published randomized clinical trials were identified using MEDLINE. Important information can be gained by adding EMBASE to MEDLINE during the search for articles, as the overlap of EMBASE and MEDLINE has been estimated to vary from 10% to 87% depending on the topic under study. Therefore, a comprehensive search strategy should include a minimum of 3 bibliographic databases (MEDLINE, EMBASE, Cochrane library), a hand-search of references of eligible studies, and direct contact with the corresponding authors of eligible articles asking for additional published or unpublished data.

   Controversy exists whether exclusion of studies in languages other than English influences the results of SR. Some authors have found no difference while others have shown that unrestricted language search-
Systematic Reviews and Meta-Analysis

This helps avoid substantial bias and increases the generalizability and applicability of the findings. Nevertheless, evidence has shown that authors are more likely to send their manuscripts to an international English-language journal if the results are positive whereas negative findings or lower methodological studies are more prone to be published in a local journal.

Publication bias is the tendency on the part of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings; i.e., studies with statistically significant and positive results are more likely to be submitted and published than articles with non-significant or negative results. Based on that, it is clear that publication bias by itself can distort the results of any SR because studies with negative or non-significant findings are more likely to be missed by the search strategy, resulting in overestimation of the pooled treatment effect.

The simplest and most common method used to detect publication bias is the funnel plot. This method plots the magnitude of the treatment effect relative to the sample size of the study. In the absence of publication bias, these plots usually have a symmetrical inverted funnel shape, because the estimates of treatment effects in smaller studies (which appear at the base of the funnel) have a larger variability compared to larger studies (which are given more weight in the pooled estimate and thus appear at the top center of the funnel) (Fig 1A). Since smaller and negative studies are less likely to be published, studies in the right bottom side of the graph are often absent, creating an asymmetrical funnel (Fig 1B). Performing a very comprehensive search strategy, as previously mentioned, can reduce publication bias.

4. Selection of Studies

It is essential throughout the selection process that decisions about inclusion or exclusion of studies be made according to a priori established criteria. The selection of studies should be conducted in such a way to minimize the risk of judgment errors (investigator bias). As these decisions often involve some degree of subjectivity, it is useful to perform these steps in duplicate, with disagreements being resolved by consensus or, when necessary, by a third reviewer.

A flow diagram, as recommend by the 2009 PRISMA statement, showing step by step how articles were selected and reasons for exclusion should be displayed in every SR.

5. Assessment of the (Methodological) Quality of Included Studies

Two reviewers, to guarantee objectivity and to avoid errors, should preferably perform the appraisal of the quality of included studies independently. As occurred for selection of studies, discrepancies between reviewers should be resolved by consensus or by consulting an independent third reviewer.

Most checklists in the literature were designed for appraisal of the quality of randomized controlled trials, but unfortunately many of these scales lack a rationale or have been used improperly. Given these issues, primary studies should not be excluded or weighted on the basis of arbitrary quality scores. Instead, quality assessment is useful for alerting reviewers and readers to the extent of bias existing in a particular review. Thus far, few comprehensive checklists for assessing quality and susceptibility to bias in observational studies have been described.

6. Data Extraction

The type of data extracted from included studies should be relevant to the review question and be pre-specified in the review protocol. Similar to article selection and quality appraisal processes, this step should be performed in duplicate, with reviewers utilizing a piloted data extraction form to safeguard against abstractor bias.

The type of outcome data extracted depends on the unit of measurement (mean, relative risk, odds ratio, or count data) that is applicable to the primary study design used in the review.

7. Synthesizing Relevant Study Results: How to Analyze and Present Results

Descriptive or non-quantitative synthesis

The objective of a descriptive or non-quantitative (qualitative) review is to correlate and present the extracted data in such a way that the characteristics (population, interventions, outcomes, study quality) and results of included studies are summarized in a meaningful way. This is best accomplished by tabulation, which allows readers to look at the evidence, its methodological rigor, and the differences between the studies.

Quantitative synthesis (meta-analysis)

Firstly, it should be determined whether quantitative synthesis is at all possible, and if so, whether it would be appropriate. Meta-analysis may not be appropriate when studies are too heterogeneous to be combined. When appropriate, the advantage of performing a meta-analysis is that pooled results can increase statistical power and lead to a more precise estimate of treatment effect when compared to the results of individual studies. Unfortunately, evidence has shown that only 45% to 68% of SR and meta-analyses have tested for heterogeneity, which may have led to a distortion of the results and spurious conclusions.

Once it is established that a meta-analysis is possible and appropriate, 3 questions should be answered a priori: Which comparisons should be made? Which outcomes should be included in the synthesis? Which effect measures should be used?

Simply calculating an arithmetic mean to summarize the results would be inappropriate, as results from small studies are more subject to random error (chance) and should be given less weight. Statistical methods used for meta-analysis employ a weighted average of the results in which the large trials are given more weight than the smaller ones.

Basically, 2 models are used to obtain this weighted average of the results: fixed and random effects models. The difference between the 2 models consists in the way the variability of the results between studies is treated. The fixed effects model assumes that the between-study variance is zero and that the true effect of treatment is the same for every study (variability exclusively due to chance) and individual studies

(continued from previous page)
Systematic Reviews and Meta-Analysis

are simply weighted by their precision (confidence interval). On the other hand, the random effects model assumes a different effect for each study and takes into account both sources of variance, within and between studies. As a result, the random effects model leads to relatively more weight being given to smaller studies and to wider confidence intervals as compared to the fixed effects model.

Rather than simply ignoring heterogeneity after applying some statistical model, one should attempt to explain it by performing a priori defined subgroup analyses. Due to the increased potential for unknown variability among observational studies, the random effects model should be used for observational data-pooling because it is more conservative.

A common criticism of meta-analyses is that some authors combine apples with oranges; i.e. they combine studies with different dose schedules, follow-ups, participants, or modes of treatment. If studies are clinically (or methodologically) too different, the results of a meta-analysis may be meaningless. In addition to this clinical heterogeneity, variability in the treatment effects of different studies is known as statistical heterogeneity.

The consideration of heterogeneity between study results is an important aspect of SR. Potential sources of heterogeneity should be defined a priori as well as plans for subgroup analyses. Heterogeneity can be assessed by several methods. Simple inspection of forest plots may be informative, as one can check for confidence intervals that do not overlap. F statistic and Cochran chi-square test (Cochran Q) - test of homogeneity - are among the most used statistical methods for assessing heterogeneity. While the Cochran Q statistic checks for heterogeneity at a predefined threshold of significance, the F statistic shows the proportion of variability across studies that can be attributed to heterogeneity rather than chance. A value lower than 25% is considered to reflect low heterogeneity, between 25% and 50% is moderate heterogeneity, and greater than 75% is high heterogeneity. There is no definitive cut-off value at which no data-pooling should be performed.

Clinical and statistical heterogeneity should be expected when pooling observational data, because confounding and selection biases often distort the findings of the primary studies. Testing for potential sources of heterogeneity may minimize these biases and generate hypotheses for future research. Therefore, heterogeneity should always be investigated, aiming to increase the scientific understanding of the studies reviewed and the clinical relevance of the conclusions drawn.

Potential sources of heterogeneity can be explained by performing subgroup analyses based on characteristics in study design, methodological quality, populations, treatments, and outcomes of the included studies. Heterogeneity can also be explored by conducting a meta-regression analysis in order to determine the effect of multiple predictor variables on the final pooled estimate.

(continued from previous page)

References

18. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. J Clin Epidemiol 2005; 58:769-776.
Health Services Research with Administrative Data and the Dartmouth Atlas of Health Care

Studies based on administrative data have been increasing in popularity among pediatric researchers. As a data source, it has many advantages. The data is already collected, it contains large numbers of patient records, and it represents patient populations across health care systems and geographic areas. Because of these attributes, administrative data is often used for health services research.

Health services research reports on how people gain access to health services, how much care costs, and the outcome of those services. The focus can be on a cohort of patients, for example, the number of children undergoing surgical intervention for vesicoureteral reflux that received a bilateral intervention. Or, a more powerful analysis is population-based research, which reports on the level of care that is received by an “at risk” population. For example, the number of children diagnosed with vesicoureteral reflux who receive a surgical intervention. However, population-based research is often difficult to perform as the denominator is the “at risk” population, which is often hard to identify.

The most often cited population-based health service research study is the Dartmouth Atlas of Health Care, originally founded by John Wennberg, which uses Medicare data as its source. Dr. Wennberg is considered to be the pioneer of variation research, and has reported on how health care is used and distributed in the United States for more than 40 years. Although the Dartmouth Atlas study’s population isn’t applicable to a pediatric audience, its findings are instrumental in understanding the forces that drive variation in health services, forces which also affect utilization of pediatric health care. The Atlas study reports utilization rates for many medical conditions. All of the rates are age, sex, race, and illness adjusted. All are per population (i.e. per 1,000 Medicare enrollees, etc.), and all are statistically significant.

One of the first things Dr. Wennberg did with the Atlas study was to define natural market areas where people sought care. Population-based studies are often reported by political boundaries, such as states or counties, which consumers cross all the time to seek care. He defined hospital referral regions (HRRs), by determining where the majority of Medicare enrollees were referred for major cardiovascular or neurological services within a Zip Code. He then assigned the Zip code to that area. This resulted in 306 HRRs across the country (Figure 1).

The Atlas found two types of variation. The first is defined as variation that is a reflection of medical need. A good example of this is hip fracture. Hip fractures are painful, everyone who has one seeks care, they are almost always correctly diagnosed, and all physicians agree on the need for hospitalization. As a consequence, the rate of hospitalization for hip fracture closely follows the actual incidence. Figure 2 shows a map of rates for hip fractures (number of hospitalizations per 1,000 Medicare enrollees) by HRR across the country. The rates are expressed as ratios to the national average, and the number of regions in each category is in parentheses. Notice that the majority of regions are within 10% above or below the national average.

The second type of variation is unwarranted variation, which is defined as variation not explained by illness, patient preference, or the dictates of evidence-based medicine. This type of variation can be seen by comparing the map of hospitalization for hip fracture to a map of hospitalization for forearm fracture, which looks vastly different (Figure 3). Compare the number of regions colored dark red, indicating a rate of 30% or higher than the national average. None of the regions for hip fracture have rates this high, while forearm fracture has 30 regions with rates 30% to three times higher than the national average.

(continued on next page)
age. Also notice that, on the forearm fracture map, regions with the highest rates in the country are often located next to a region with the lowest rate in the country – there is no discernable pattern to level of hospitalization for this type of fracture.

The Atlas identified three categories of unwarranted variation. The first is variation in the use of effective care and in outcomes of care such as surgical mortality rates. Variation in this category is usually attributed to medical error, which is a product of system error. Sometimes medical errors result from the failure to provide specific treatments that, according to good medical science, should be provided to all eligible patients. For example, the use of beta blockers following heart attacks. Figure 4 shows the distribution of “ideal” patients receiving beta blockers at time of discharge among the 306 hospital referral regions. Each dot is a HRR. The “right rate” dictated by medical science is 100%. In every region, there was underuse of beta blockers. In only eight regions did 80% or more of patients receive beta blockers upon discharge from the hospital. In the regions with the worst compliance with practice guidelines, only 10% of patients for whom beta blockers would have been appropriate received them.

The Atlas study proposes that the remedy for reducing variation in the use of effective care involves the redesign of practice. It is a systems error. For example, certain HMOs have high rates of immunizations. They believe that this is, in part, due to organizational efforts by HMOs to ensure that effective services are performed. HMOs with good records in providing effective care don’t depend on the physician’s memory to assure that care is given. They depend on systems of care that ensure compliance.

The second category of unwarranted variation identified by the Atlas is variation that reflects the misuse of discretionary care. Some conditions can be treated in more than one medically valid way, usually a choice between a medical and a surgical approach. In these situations there is no fixed, medically correct way that is right for every patient. Choosing the right treatment should depend on the patient’s own attitude toward the risks and benefits of the alternatives.

Often, these variations can be attributed to lack of evidence - the outcomes of the alternatives are simply not known. One example used by the Atlas is the treatment of early stage prostate cancer. Variation in rates of prostate surgery reflects differences in recommendations of local physicians regarding the screening and treatment of prostate cancer. Medical opinion is deeply divided concerning the value of aggressive treatment of early stage prostate cancer. Different specialists have different theories about which treatment works best, but the clinical studies necessary to test these theories haven’t been performed. Radiologists are more likely to recommend radiation treatment, urologists are more likely to recommend surgery, and some physicians recommend watchful waiting, especially for older men. Therefore, treatment of prostate cancer is strongly influenced by the lack of evidence-based information on outcomes.

This variation can easily be seen in Figure 5, which shows two maps of Florida; one with rates of prostate cancer surgery and one with rates of colorectal cancer surgery. Again, both maps report the ratio of rates to the national average. Unlike early stage prostate cancer, outcomes of different treatment options are better established for colorectal cancer. Thus, the rates of colectomy are much more homogenous. While rates for prostate cancer surgery varied from 75% below the national average to almost three times the national average, only a few regions had rates more than 25% below the national average, and no Florida HRR had rates of colon cancer surgery as much as 30% higher than the national average.

The Atlas study proposes that the remedy for reducing variation in discretionary care due to poor quality of clinical science is outcomes research. And when clinical science is present in conditions which can be treated in more than one medically valid way, the remedy for reducing variation is shared decision making, the active involvement of the patient in the choice of discretionary care.

Finally, the third category of variation is in what the Atlas calls supply-sensitive services. Unlike variation in discretionary treatment, where strongly held medical opinions about the value of one form of treatment over another drive clinical decision making, decisions influencing the frequency of use of supply-sensitive services is not supported by well articulated medical ideas, much less by clinical evidence.

This variation is associated with a region’s supply of health services, such as the number of physicians or hospital beds, and includes rates of diagnostic testing, physician visits, and hospitalizations. Practice style,
Health Services Research

(continued from previous page)

such as frequency of follow-up visits, when to refer to a specialist, or when to hospitalize a patient, are not subjects of most medical texts. In fact, the Atlas study found most medical conditions to be supply-sensitive. Figure 6 shows the relationship between hospital beds per 1,000 Medicare residents and rates of hospitalization for medical services (all non-surgical services combined). The $R^2$ of the association between hospital beds per 1,000 residents and rates of hospitalization for medical conditions is 0.54, meaning that hospital capacity explains over half of the variation in rates of hospitalization for all medical conditions. In contrast, there is almost no association between hospital bed supply and hospitalization for hip fracture.

The Atlas theorized that patterns in the use of supply-sensitive services appear to reflect an assumption that more health care produces better health. In other words, variation is driven by an assumption that if healthcare resources are available, they should be used to treat the sick, which in turn leads to better outcomes. Because of this assumption, a health care system’s capacity highly influences utilization.

It is easy to see why this should be the case. For example, take the subject of visits to cardiologists. Most visits are scheduled follow-up visits. But, how long should the interval between follow-up visits be? There are no set guidelines addressing this question. It’s only common sense that, in a region with twice as many cardiologists per 1,000 Medicare enrollees, there will be twice as many office hours available to fill. And, in the absence of guidelines and under the assumption that more is better, scheduling of most follow up visits is highly influenced by availability. Figure 7, depicting the association between number of cardiologist per 1,000 Medicare enrollees and number of visits to cardiologists, shows that almost half of the variation in visits is driven by supply. The Atlas found that services for most medical conditions are highly influenced by system capacity - it is supply driven - which lead them to conclude that “geography is destiny”.

This finding lead the Atlas study to ask, is more better? One of the first places the Atlas looked to answer that question was in end of life care – that is, the levels of care the Medicare population received in their last six months of life. This is an unusual cohort in that the outcome is known. At the end of six months, they were all deceased. The Atlas found striking levels of variation across all areas of end of life supply sensitive services. For example, Figure 8 depicts the number of physician visits per decedent in their last six months of life. Assuming one clinic visit per day, in the regions with the highest number of physician visits, Medicare enrollees spent 30 to 60 days of their last six months of life attending a clinic visit. In areas with the lowest numbers, enrollees spent 15 to 20 days visiting a doctor’s office. If they lived in New Brunswick, NJ, Newark, NJ, or Los Angeles, CA, they averaged almost 60 visits to a physician. If they lived in Idaho Falls, ID, Mason City, IA or Billings, MT, they averaged about 30 visits to a physician. The enormity of the variation seen between these regions illustrates the impact that supply can have on not only the expense of the health care system as a whole, but also on the quality of life at an individual level.

From their end of life research, The Atlas concluded that, if more conservative practice styles do not appear to result in a trade-off between the length and the quality of life, then providing fewer supply-sensitive services does not mean health care rationing. Rather, the health care systems serving regions with low rates of supply-sensitive services are benchmarks for more efficient patterns of resource allocation and spending. And that the remedy for reducing variation in supply-sensitive services is to more effectively manage capacity and to promote conservative practice patterns.

So, is more better? Based on their study of the end-of-life cohort and other cohorts, Dartmouth says no. And not only is it not better, it’s very expensive. Dartmouth Atlas study team members estimate that approximately 30% of care in the U.S. is unnecessary. A recent study estimated that unnecessary care added an additional $830 billion to the cost of health care in 2009 alone. This waste is due to several factors, including overdiagnosis, inadequate scientific knowledge, oversupply of health care infrastructure, and non-adherence to existing evidence-based guidelines. Not only are unnecessary care, overdiagnosis, and overtreatment expensive, they are widely reported to increase the risk of harm and potentially lead to worse outcomes.

What do the Atlas findings mean to the pediatric urologic community? For one thing, we are operating with a health care system that many agree is broken, and that is quickly going bankrupt. Health care costs are rising at an insupportable rate. Children have more to gain and provide larger cost savings than adults by proper management of disease, as there are more life years at stake. Little empirical evidence of outcomes based on strong clinical science exists for many pediatric treatment options.

However, claims data often cannot answer a very important question, “Which rate is right?” Even rigorous studies such as the Dartmouth Atlas of Health Care can rarely answer this question, as the answer lies in comparative effectiveness research. Claims data lacks sufficient granularity to discern severity of illness or results of diagnostic testing. As it is gathered for billing purposes, not research purposes, it’s limited to billing codes, ICD-9 diagnosis and procedure codes or CPT codes. As such, outcomes often cannot be teased out. It’s mainly inpatient data only, which excludes the study of many conditions that are treated in an outpatient setting.

(continued on next page)
Health Services Research

More pediatric claims data sources have become available in recent years. The Pediatric Health Information Systems (PHIS) database contains data from 43 free-standing children’s hospitals across the country, although your hospital must participate in PHIS for you to have access to the data. The KIDs inpatient database, available to everyone, is a 20% sample of national data including not only children’s hospitals, but also community hospitals. In addition, many states mandate data reporting which is available to researchers. However, it should be noted that none of these data sources report the “at risk” population (denominator), which limits their use in health services research to utilization of types of interventions rather than the levels of interventions that a population of children are receiving.

In spite of the limitations inherent in most widely available administrative pediatric data sources, health services research, a very important starting point to improving health care, is possible. Research on variation in the types of interventions enables the identification of benchmark areas. It identifies conditions with high or low variation of interventions. It promotes a dialogue. What are the physicians in high utilization areas doing that is different from physicians in low utilization areas? Are their self-reported outcomes the same? What is the difference in cost for these treatment options? In turn, this dialogue can lead to multi-centered clinical studies to obtain the scientific evidence needed to determine which treatment works best or which rate is right. And from strong scientific evidence, consensus statements and guidelines can be formulated, leading to a more effective, less costly delivery of health care.

References

Concomitant Ovotesticular DSD and Congenital Adrenal Hyperplasia

William P. Parker, MD1, John M. Gatti, MD2, J. Patrick Murphy, MD2
1Department of Urology University of Kansas Medical Center, Kansas City, Kansas, 2Department of Surgery and Urology, Children’s Mercy Hospital, Kansas City, Missouri

Introduction
Genital ambiguity occurs in approximately 1 in 4500 patients1 with causes ranging from disorders of gonadal differentiation to disorders of steroid biosynthesis. Ovotesticular DSD represents approximately 3-10% of all cases of DSD.2 Additionally, CAH is a well-known cause of ambiguous genitalia in the neonate occurring in 1 in 1800 newborn children.3 We report a case of concomitant ovotesticular DSD and congenital adrenal hyperplasia.

Case Report
A 1 day-old term infant born to a G1P1 female with choorioamnionitis was referred to our institution for the evaluation of genital ambiguity noted at birth. The patient had a normal physical findings with the exception of the genital exam. The patient had a stretched phallic length of 2.5cm with significant ventral chordee. The urogenital meatus was located in the fused labioscrotal folds which failed to reveal palpable gonads.

Radiologic evaluation included a pelvic ultrasound and a MRI of the abdomen and pelvis. The pelvic ultrasound demonstrated the presence of a unicorne uterus and no abnormalities of the upper tracts. On MRI the patient was noted to have a well-defined uterus with a short vagina leading to the confluence of the proximal urethra, forming a common urogenital sinus. Gonads were visible within the abdomen. Laboratory evaluation revealed normal serum electrolytes and elevated 17-hydroxyprogesterone (2900ng/dl; nml 3-90ng/dl) and testosterone (800ng/dl; nml 72-344ng/dl) levels. Karyotype analysis was positive for 45XO (75%), 46XY(t8) (12.5%), and 46XY(t16) (12.5%). Genomic evaluation confirmed a deletion in the 21-hydroxylase region of chromosome 6.

The patient was subsequently taken to the operating room where a cystoscopy and laparoscopy were performed. Cystoscopy was notable for a 1.0cm length from the urogenital meatus to the urethrovaginal confluence and 1.5cm from the confluence to the bladder neck. The vagina was of adequate length with a single cervix at the apex. Laparoscopy revealed bilateral gonads associated with bilateral fallopian tubes and the absence of vas deferens bilaterally. Due to the presence of 46XY mosaic gonadectomy was performed. Final pathology was significant for a left ovotestis and a right ovary.

Discussion
Ovotesticular DSD is the rare combination of both ovarian and testicular tissue in the same patient.4 Genetically most patients (65%) with ovotesticular DSD have a 46XX karyotype with 25% containing a XY mosaicism.5 Of those with XY mosaicism, the 45XO, 46XY mosaic is rarely reported in the literature.2 In the patient with ovotesticular DSD the degree of external genitalia virilization is related to the amount of androgen production in the dysgenic testicular tissue. In the present case, the phenotypic presentation is further amplified by the presence of CAH, which is a well-known and described cause of excess androgen production and external genitalia virilization. To our knowledge this is the first case of concomitant ovotesticular DSD and CAH.

References
A Conservative Approach to the Ectopic Ureterocele

Heather McCaffrey, M.D.1, Thomas Forest, M.D.2, Robert Wasnick, M.D.1
1Stony Brook University Medical Center, Stony Brook, NY, 2Pediatric Urology Clinic, Lafayette, LA

Introduction

There has been an evolving trend towards a more conservative approach in the management of renal duplication anomalies associated with ureterocele, especially in those detected antenatally. When indicated, endoscopic incision of the ureterocele is often the initial treatment choice, especially if an obstructive component is identified. More invasive surgical intervention may be necessary in the presence of high-grade reflux, progressive hydronephrosis, and/or recurrent urinary tract infection. Conversely, “watchful waiting” is a reasonable option if the upper tracts remain stable and the clinical course benign. The authors describe a case of a four year-old female patient with a unilateral duplex system and ureterocele detected postnatally in which conservative management was successful.

Case Presentation

A four year-old female incurred a febrile urinary tract infection (UTI) at one year of age, which resolved with oral antibiotic treatment. Renal ultrasound demonstrated a left duplication anomaly, upper moiety hydroureteronephrosis and a one cm left sided ureterocele (Fig 1). There was excellent bladder emptying and no vesicoureteral reflux on voiding cystourethrography (VCUG). After completing treatment for the UTI she was placed on low dose antibiotic prophylaxis and did well clinically. Endoscopic evaluation at 16 months of age revealed an ectopic ureterocele at the bladder neck, which appeared non-obstructive readily admitting a 7.6 French cystoureteroscope. Retrograde ureteropyelography demonstrated mild dilation of the ureter draining the lower moiety and marked dilation and tortuosity of the ureter draining the upper moiety (Fig 2). The upper pole renal pelvis was dilated and there was poor drainage of this moiety. Conservative management and surgical intervention were presented to the parents who opted for surveillance with serial renal ultrasonography. A short course of antibiotic prophylaxis was continued after the cystoscopy and subsequently stopped after three months. Serial ultrasonography over a three-year period has shown an unchanged appearance of the upper tracts with fixed dilation of the upper moiety collecting system and ureter (Fig 3). The patient has remained infection-free and has complete daytime continence with variable nocturnal enuresis.

Figure 1: Initial ultrasound of the left kidney and bladder demonstrating a duplication anomaly and left ureterocele at one year of age. The right kidney was normal.

Figure 2: Retrograde ureteropyelography demonstrating mild lower pole ureteral dilation (left image) and severe left upper moiety hydroureteronephrosis (right image).

Figure 3: Left renal ultrasound demonstrating stable upper pole caliectasis in this four-year old patient.

Conclusion

The authors describe a conservative approach to the ectopic ureterocele resulting in an acceptable clinical outcome in a four-year old female. This approach can be a valid option in patients with a renal duplication anomaly if both moieties of the duplicated system are not obstructed and there is no evidence of bladder outlet obstruction or vesicoureteral reflux. These patients must also stay free of infection to avoid renal parenchymal scarring. Management strategies must be tailored to fit the individual patient and clinical scenario.

References


Dialogues in Pediatric Urology

February, 2013

Cloacal Exstrophy in Polyzygotic Multiples: A Case Series

Christopher L Powell 1, John M Gatti 2, J Patrick Murphy 2

1Department of Urology University of Kansas Hospital, Kansas City, Kansas
2Department of Surgery and Urology, Children’s Mercy Hospital, Kansas City, Missouri

Introduction

OEIS complex (omphalocele-exstrophy-imperforate anus-spinal defects) or cloacal exstrophy (CE) is a rare anomaly representing a distinct presentation of the exstrophy-epispadias complex. Incidence of CE is estimated to be between 1:200,000 and 1:400,000 live births. Recent literature has reported a number of monozygotic twins presenting with CE. Rarely has CE been reported in multiples of polyzygotic births resulting in a single case of CE and viability of all sibling neonates. We present an unusual case series of three such instances.

Case 1

This male patient presented following cesarean section delivery at 38 weeks gestation. He was a dizygotic multiple. Upon evaluation the patient was found to have cloacal exstrophy, omphalocele, imperforate anus, duplicate colon, left cryptorchidism, right anorchidism, and myelomeningocele of the sacral, lumbar, and thoracic areas. On day three of life, he underwent separation of the colon from the bladder, tubularization of the cecum, creation of a single end colostomy using both colonic limbs, left orchiopexy, bladder plate combination, and omphalocele closure. At 10 months of age he underwent pelvic osteotomies with external fixation and bladder exstrophy closure. At a three-month post-operative encounter he continued to do well.

Case 2

This male patient presented following delivery at an outside hospital. He was a dizygotic multiple. Upon evaluation the patient was found to have cloacal exstrophy, omphalocele, imperforate anus, atrial septal defect, patent foramen ovale, lumbar myelomeningocele, and bilateral dysplastic kidneys. Despite aggressive fluid resuscitation his creatinine continued to rise and the patient expired on day six of life.

Case 3

This male patient presented following cesarean section delivery at 30 weeks gestation. He was a polyzygotic multiple. Upon evaluation the patient was found to have cloacal exstrophy, omphalocele, imperforate anus, duplicate appendix, and a separate orifice likely associated with the hindgut. He was also found to have a tethered cord with sacral agenesis and a patent foramen ovale. On day nine of life, he underwent separation of the colon from the bladder, tubularization of the cecum, creation of an end colostomy, bladder plate combination, and omphalocele closure. He is currently awaiting bladder exstrophy closure.

Discussion

The historical incidence of cloacal exstrophy is 1:200,000 and 1:400,000 live births. Incidence may be increasing as recent studies have indicated rates of 146, 000 to 184,195 . Others have postulated a decreasing incidence due to increasing early diagnosis and elective termination. Despite survival rates that now approach 100%, the pathogenesis of CE continues to be poorly understood. Multiple reports of CE in monozygotic and discordant dizygotic twins suggest an underlying genetic link. Reports of CE amongst non-multiple family members have involved multi-generational gaps and a genetic etiology was not identified. Others have suggested environmental factors including maternal exposure to clomiphene citrate. A recent study reported CE rates in 24 countries resulting in 186 cases. 18 of 186 (9.7%) were products of a multiples producing gestation. Four of the 18 (22%) were dizygotic twins. Our series highlights an unusually high prevalence of polyzygotic multiples with CE originating from a single hospital’s referral zone.

References

Ehlers-Danlos Syndrome with Posterior Urethral Valves: For Better or Worse

Christopher L Powell 1, John M Gatti 2, J Patrick Murphy 2
1Department of Urology University of Kansas Hospital, Kansas City, Kansas
2Department of Surgery and Urology, Children’s Mercy Hospital, Kansas City, Missouri

Introduction

Ehlers-Danlos syndrome (EDS) or Cutis hyperelastica is a group of rare, inherited connective tissues disorders characterized by a defect in the synthesis of collagen. Reported rates have increased over time and it is now thought to be present in 1:5,000 births to some degree. Increased elasticity of connective tissue often results in genitourinary complications such as bladder diverticula and impaired detrusor function. We present a previously undocumented combination and unexpected presentation of EDS in a patient with previously diagnosed posterior urethral valves (PUV).

Case Report

This patient was the product of a 38-week gestation. He had antenatal studies which demonstrated dilated hydronephrosis and therefore postpartum imaging was obtained. He had an ultrasound on day of life two showing bilateral SFU grade III hydronephrosis. His renal function remained within normal limits. A cystogram was performed on day of life three which demonstrated a relatively smooth bladder wall with mild trabeculation. There were diverticula at each bladder base likely representing paraureteral diverticula but there was no evidence of vesicoureteral reflux. During the voiding images the prosstatic fossa was mildly dilated and minimal bladder neck hypertrophy was identified. On day of life twelve cystourethroscopy revealed type 3 posterior urethral valves. The valves were ablated revealing a mildly trabeculated bladder and paraureteral diverticula. As he aged, he developed persistent daytime and nocturnal enuresis and was closely followed for voiding dysfunction with constipation. Urodynamics at six years of age revealed normal bladder capacity, compliance, and contractility with minimal detrusor sphincter dysynergia and detrusor overactivity. Post-void residuals were within normal limits. He was treated with an alpha-blocker for detrusor sphincter dysynergia but symptoms persisted and he was transitioned to anti-cholinergic therapy for detrusor overactivity. By nine years of age his incontinence episodes resolved. His sister was recently diagnosed with Ehlers-Danlos syndrome, hypermobility type (type 3). Upon genetic evaluation, this patient was also found to have autosomal dominant Ehlers-Danlos syndrome. His renal function remains within normal limits.

Discussion

Ehlers-Danlos syndrome represents an increasingly diagnosed collection of inherited connective tissue disorders characterized by a defect in the synthesis of collagen. Current practice organizes EDS into 6 categories based on genetic mutations and clinical symptoms. These include the following: hypermobility, classical, vascular, kyphoscoliosis, arthrochalasis, and dermatosparaxis. Posterior urethral valves are theorized to be caused by abnormal migration of mesonephric ducts with resulting midline fusion. Posterior urethral valves occur in 1 in 8,000 to 25,000 live births. Multiple cases of bladder diverticula in the setting of EDS have been reported. Diverticula in this setting are thought to be associated with worse outcomes. Case studies have previously described bladder outlet obstruction secondary to an enlarged diverticulum in a patient with EDS. EDS while previously associated with worse genitourinary outcomes may have been protective in this patient with PUV due to an improved pop-off mechanism. To our knowledge, this is the first patient reported with EDS and concurrent PUV.

References

All Posterior Urethral Valves Are Not Alike

Heather McCaffrey, M.D.¹, Thomas Forest, M.D.², Robert Wasnick, M.D.¹
¹Stony Brook University Medical Center, Stony Brook, NY, ²Pediatric Urology Clinic, Lafayette, LA

Introduction
Poster urethral valves can present with clinical variability, yet tend to have a classic constellation of findings. The authors describe a neonate with posterior urethral valves (PUVs) who had a somewhat atypical presentation and clinical course.

Case Presentation
The patient was initially evaluated and managed at a nearby institution and was antenatally found to have bilateral hydronephrosis. The expectant mother subsequently developed sudden and severe oligohydramnios at 30 weeks gestational age (GA), resulting in induction and early delivery. Initial postnatal renal and bladder ultrasonography demonstrated severe bilateral hydronephrosis (Figure 1) and a distended, thick walled bladder (Figure 2). Voiding cystourethrography (VCUG) revealed a mildly dilated posterior urethra and bilateral grade V vesicoureteral reflux (VUR), consistent with posterior urethral valves (Figure 3). The patient was initially treated with indwelling catheter drainage of his bladder while in the neonatal intensive care unit. He resumed spontaneous voiding upon catheter removal and serial renal ultrasonography showed an improved degree of hydronephrosis. His serum creatinine on discharge from this institution was 0.7 mg/dL.

The patient underwent endoscopic evaluation at two months of age after being referred to the author’s institution. Cystourethroscopy revealed a minimally trabeculated bladder with both ureteral orifices in normal anatomic position. Type one PUVs emanating from the verumontanum and coapting only slightly on the dorsal aspect of the urethra were visualized. This configuration appeared only minimally obstructive. The patient underwent endoscopic fulguration using a handheld electrocautery unit and a stylet passed through a 5 French open-ended ureteral catheter. Post-procedure expression-induced voiding demonstrated a generous urinary stream. His nadir serum creatinine obtained the day of surgery was 0.3 mg/dL. His is currently voiding well with an improved appearance of his kidneys on renal ultrasound (Figure 4).

Discussion
The typical characteristics of PUVs on antenatal ultrasonography include marked hydronephrosis with a distended and/or thickened bladder wall. VCUG remains the most important study for diagnosis. Once diagnosed, bladder drainage followed by endoscopic ablation remains the standard of care. This case is somewhat aberrant as endoscopic ablation was delayed two months for both logistic and clinical reasons. Serial renal ultrasonography demonstrated improved hydronephrosis while serum creatinine level improved with spontaneous voiding. This clinical course would not be expected in the typical PUV patient. PUVs may represent a spectrum of obstructive uropathy with a minimally obstructive presentation in the patient described.

References
Fibroepithelial Polyp of the Verumontanum Presenting with Bladder Outlet Obstruction

Introduction

We present an unusual case of a fibroepithelial polyp of the verumontanum. This originally presented with a prenatal ultrasound demonstrating unilateral hydronephrosis and bladder wall thickening.

Case Report

A 3-week-old male presented to clinic after being followed prenatally for mild left-sided hydronephrosis and bladder wall thickening. Although the mother had normal amniotic fluid levels throughout pregnancy, the baby was delivered via cesarean section at 39 weeks secondary to oligohydramnios. Physical examination revealed a normal genitourinary exam including a circumcised phallus, patent orthotopic meatus, bilateral descended testes, and normal back and abdominal exams free of lesions or masses.

Renal ultrasound at three weeks of age demonstrated left-sided SFU Grade 2 hydronephrosis without hydroureter and a normal right kidney. The bladder was thick walled with a polypoid solid lesion measuring 5mm near the base of the bladder. No dilated posterior urethra was visible on this ultrasound examination.

The patient underwent cystoscopy, which revealed no valvular tissue in the posterior urethra. We visualized a pedunculated polypoid lesion arising directly from the superior aspect of the verumontanum. Cystoscopic examination of the bladder revealed mild trabeculations and an edematous bladder neck, but there was no erythema of the urothelium, masses, or intravesicular lesions. Cup biopsy forceps were used to remove the tan, rubbery lesion, which was confirmed by pathology to be a 0.6cm benign fibroepithelial polyp.

Discussion

Though rare, lower urinary tract benign fibroepithelial polyps have been reported from the middle calyx of the kidney to the anterior urethra. They occur most commonly in infancy and childhood, and can present with dysuria, hematuria, straining to void, urinary tract infections, obstruction, and urinary retention.

The etiology fibroepithelial polyps remains unclear. It has been hypothesized that urethral polyps may result from a defective protrusion of the urethral wall, effects of maternal estrogen, or congenital anomalies. Histopathologic evaluation of fibroepithelial polyps reveals a connective tissue and smooth muscle core surrounded by normal transitional epithelium. No cases of malignant transformation of a fibroepithelial polyp have been reported. The polyps do not seem to recur, and thus follow-up cystoscopies and cytologies do not appear to be necessary.

References

A Case of Concurrent Prune Belly Syndrome and Posterior Urethral Valves Leading to Persistent Renal Failure Despite Maximal Upper Tract Drainage

Stephanie Rubino MD, Fernando Ferrer MD, Cynthia Silva MD, Christina Kim MD
University of Connecticut, Connecticut Children’s Medical Center

Introduction

We present an unusual case of an infant with both prune belly syndrome (PBS) and posterior urethral valves (PUV) with bilateral SFU Grade 4 hydronephrosis. This patient developed acute renal failure and oliguria requiring not only cutaneous vesicostomy but also a subsequent cutaneous ureterostomy for upper tract drainage.

Case Report

At 20 weeks gestation, this male fetus was found to have a left multicystic dysplastic kidney and right severe hydroureteronephrosis on screening ultrasound. At 29 weeks and 4 days there was anhydramnios and the baby was delivered by cesarean section.

Upon birth, physical examination revealed a softly distended, floppy abdomen with visible loops of bowel. Genitourinary exam revealed a normal uncircumcised phallus with a patent orthotopic meatus, bilateral non-palpable testes and an orthotopic anus.

Renal ultrasound after birth revealed bilateral cystic appearing kidneys with right hydrourerteronephrosis. Voiding cystourethrogram (VCUG) performed 24 hours after birth demonstrated right-sided vesicoureteral reflux, a markedly contracted bladder, and a significantly dilated posterior urethra suggestive of posterior urethral valves. The patient also had an elevated serum creatinine with nadir at a level of 2mg/dL with urinary catheter drainage.

The patient was subsequently brought to the operating room for cystoscopy with valve resection. Despite ablation, he continued to have high post-void residuals and required intermittent catheterization. After a short nadir, his creatinine remained elevated, so we proceeded with a cutaneous vesicostomy.

After vesicostomy, the patient remained in renal failure with right hydroureteronephrosis and periodic oliguria. As a last attempt to improve his creatinine, we proceeded with right-sided cutaneous loop ureterostomy for proximal diversion. The right ureter was very dilated and tortuous but without any obvious proximal or distal obstruction. Later, the creatinine nadired to 2.4mg/dL. Follow up renal ultrasounds one month and two months after ureterostomy were unchanged from prior exams.

Discussion

Findings of Prune Belly Syndrome (PBS) include renal dysplasia, hydroureteronephrosis, and an enlarged bladder. These findings may be related to histologic changes including mesenchymal defects\(^1\) and an increased ratio of collagen to muscle fibers\(^2\) leading to poor peristalsis, or may be secondary to outlet obstruction.

Posterior urethral valve tissue manifests in three different types with various etiologies.

This obstructing tissue results in high-pressure storage and voiding of urine, resulting in detrusor hypertrophy, increased collagen, and thus loss of compliance as well as detrusor hyperreflexia. This promotes hydroureteronephrosis, VUR, and resulting renal damage.

Both PBS and PUV promote progressive renal deterioration; the former by incomplete emptying and the latter by bladder outlet obstruction. Each patient will benefit from urinary storage at safe pressures, complete emptying, and effective bladder cycling. This patient was confusing with a persistent creatinine rise despite maximal drainage of his bladder.

References