Τα λάθη στο δείγμα και τη στατιστική ανάλυση

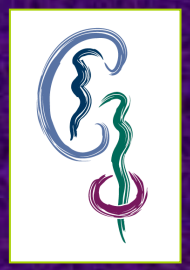
Χαράλαμπος Κ. Μαμουλάκης

Επικουρος Καθηγητής Ουρολογίας

Ουρολογική Κλινική

Πανεπιστημιακό Γενικό Νοσοκομείο Ηρακλείου

Πανεπιστήμιο Κρήτης, Τμήμα Ιατρικής



8ι "#$%&$'(#)* +, !-./! 01&!-!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Σύγκρουση Συμφερόντων: Καμία

8! "#$%&$ '(#) +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr*

Πώς ερμηνεύεονται το αποτέλεσμα μιας μελέτης

- Μπορεί να είναι συνέπεια:

  - Αλήθειας, δηλαδή ισχύει (True)

  -Μεροληψίας (bias)
  δηλαδή συστηματικού σφάλματος (systematic error)

  - Τύχης, δηλαδή τυχαίου σφάλματος (random error)

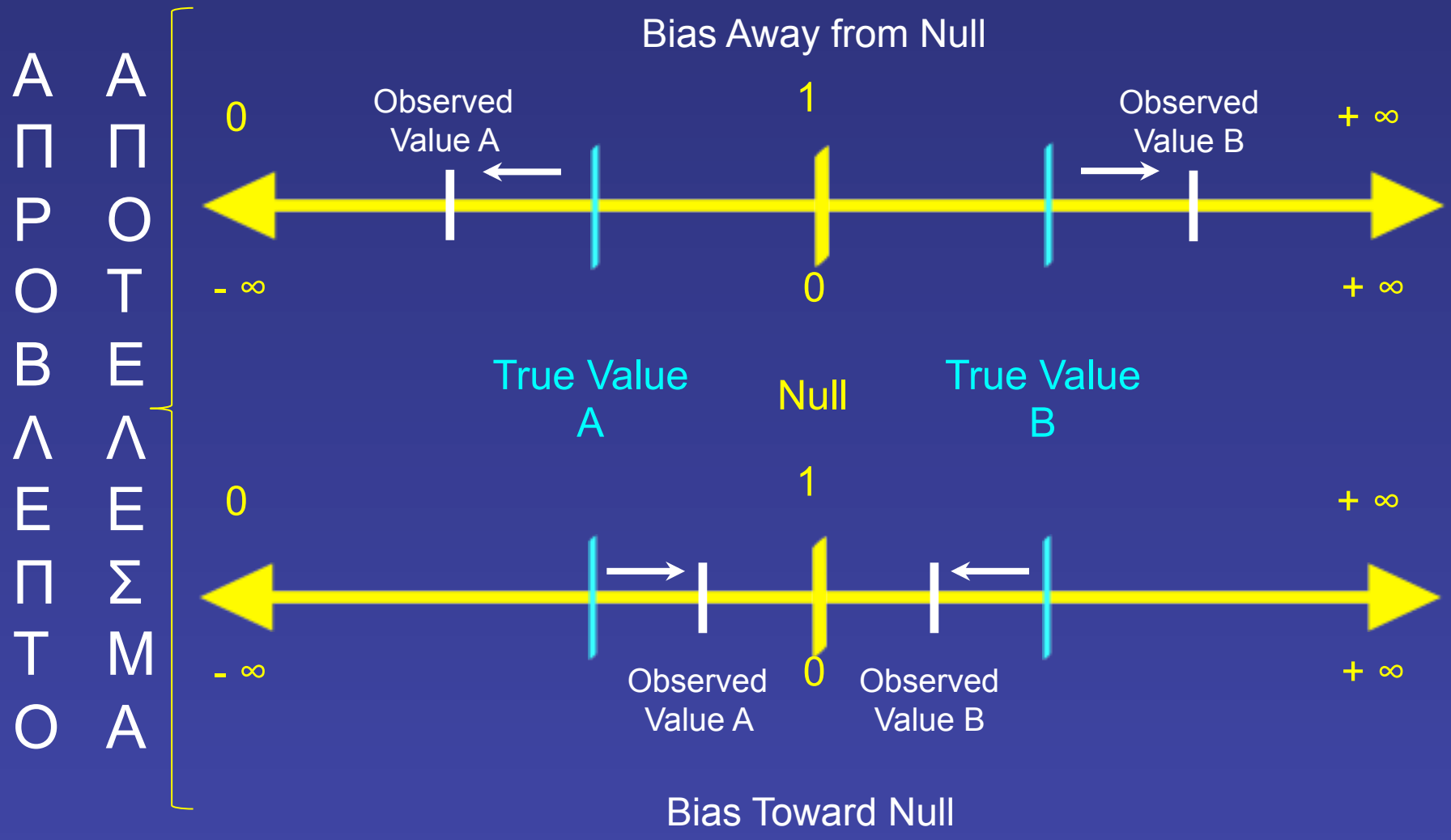
Συστηματικά (systematic) vs.Τυχαία (random) σφάλματα

- Systematic error (bias):

  3 μορφές (για όλους τους τύπους σχεδιασμού μελετών)
  Κάθε σχεδιασμός διαθέτει τρόπους *a priori* ελέγχου των σφαλμάτων από την εφαρμογή των οποίων εξαρτάται ο βαθμός εγκυρότητας της μελέτης
  - Selection bias
  - Confounding
  - Information bias (measurement error)

- Random error (statistical precision)
  - Type I errors
  - Type I errors

Τα συστηματικά σφάλματα είναι σοβαρά και προκύπτουν από πλημμελή αρχικό σχεδιασμό μιας μελέτης



Τα συστηματικά σφάλματα κατά κανόνα δεν μπορούν να ελεγχθούν εκ των υστέρων (κατά την στατιστική ανάλυση)

Συστηματικό Σφάλμα:
Σφάλμα Επιλογής (Selection bias)

8' "#$%&$ '(#)* +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Ο πληθυσμός της μελέτης είναι αντιπροσωπευτικός του γενικού πληθυσμού;

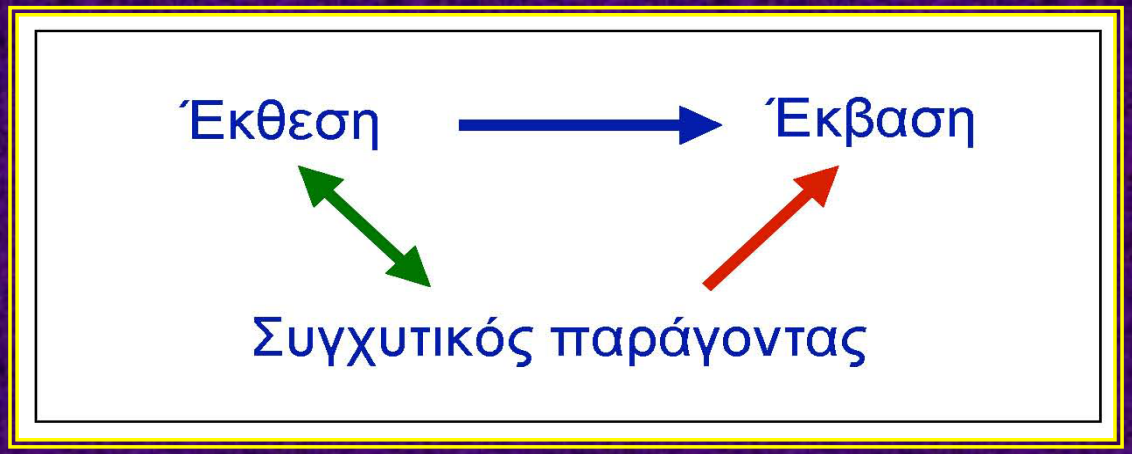
Η απόφαση των ασθενών για το αν θα συμμετέχουν (ή θα παραμείνουν) στη μελέτη εξαρτάται από τη βαρύτητα της νόσου τους;

Το σφάλμα επιλογής περιορίζεται με προσεκτικό σχεδιασμό των κριτηρίων εισόδου/αποκλεισμού

Συστηματικό Σφάλμα: Φαινόμενο Σύγχυσης (Confounding)

8ι "#$%&$ '(#)* +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Ο «συγχυτικός» παράγοντας έχει 2 βασικά χαρακτηριστικά:



- Σχετίζεται με την έκθεση χωρίς να είναι αποτέλεσμα της
- Σχετίζεται με την έκβαση ανεξάρτητα των επίπεδων έκθεσης

Smoking

Yellow Finger

Lung Cancer

8ι "#$%&$ '(#)* +, ! -./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Τρόποι ελέγχος του φαινομένου σύγχυσης

- RCT:
  - Τυχαιοποίηση (εφόσον είναι επιτυχής)
  - Ανάλυση με πρόθεση θεραπείας (ITT analysis)
  - (ή έστω) εφαρμογή μοντέλων παλινδρόμησης σε περιπτώσεις πολύ ασύμμετρων απωλειών στις ομάδες

- Μελέτες Παρατήρησης (Observational Studies)
  - Περιορισμός (Restriction) κριτηρίων εισόδου/αποκλεισμού (περιορισμός πληθυσμού σε πιο ομοιγενείς ομάδες)
  - Στατιστική ανάλυση (έλεγχος συγχυτικών παραγόντων)
    - Διαστρωμάτωση (Stratification)
    - Μοντέλα παλινδρόμησης (Regression Models)
    - Αντιστοίχηση (Matching)

ΠΡΑΚΤΙΚΟ ΠΑΡΑΔΕΙΓΜΑ
(Τυχαιοποιημένη Ελεγχόμενη Κλινική Δοκιμή)
(Rrandomized Controlled Trial)

8! "#$%&$ '(#)* +, ! -./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Platinum Priority – Benign Prostatic Obstruction

Editorial by Alexander Bachmann et al. on pp. 677–679 of this issue

Midterm Results from an International Multicentre Randomised Controlled Trial Comparing Bipolar with Monopolar Transurethral Resection of the Prostate

Charalampos Mamoulakis [a,b,], Michael Schulze [c], Andreas Skolarikos [d], Gerasimos Alivizatos [d], Roberto M. Scarpa [e], Jens J. Rassweiler [c], Jean J.M.C.H. de la Rosette [a], Cesare M. Scoffone [e]*

Mamoulakis et al. Eur Urol. 2013;63:667-676

8! "#$%&$ '(#) +, ! -./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr*

Background: Pooled data from randomised controlled trials (RCTs) with short-term follow-up have shown a safety advantage for bipolar transurethral resection of the prostate (B-TURP) compared with monopolar TURP (M-TURP). However, RCTs with follow-up >12 mo are scarce.

Objective: To compare the midterm safety/efficacy of B-TURP versus M-TURP.

Design, setting, and participants: From July 2006 to June 2009, TURP candidates with benign prostatic obstruction were consecutively recruited in four centres, randomised 1:1 into the M-TURP or the B-TURP arm and regularly followed up to 36 mo postoperatively. A total of 295 patients were enrolled.

Intervention: M-TURP or B-TURP using the AUTOCON II 400 electrosurgical unit.

Outcome measurements and statistical analysis: Safety was estimated by complication rates with a special emphasis on urethral strictures (US) and bladder neck contractures (BNCs) recorded during the short-term (up to 12 mo) and midterm (up to 36 mo) follow-up. Efficacy quantified by changes in maximum urine flow rate, postvoid residual urine volume, and International Prostate Symptom Score was compared with baseline, and reintervention rates in each arm were also evaluated.

Results and limitations: A total of 279 patients received treatment after allocation. Mean follow-up was 28.8 mo. A total of 186 of 279 patients (66.7%) completed the 36-mo follow-up. Posttreatment withdrawal rates did not differ significantly between arms. Safety was assessed in 230 patients (82.4%) at a mean follow-up of 33.4 mo. Ten US cases were seen in each arm (M-TURP vs B-TURP: 9.3% vs 8.2%; $p = 0.959$); two versus eight BNC cases (M-TURP vs B-TURP: 1.9% vs 6.6%; $p = 0.108$) were collectively detected at the midterm follow-up. Resection type was not a significant predictor of the risk of US/BNC formation. Efficacy was similar between arms and durable. A total of 10 of 230 patients (4.3%) experienced failure to cure and needed reintervention without significant differences between arms. High overall reintervention rates, withdrawal rates, and sample size determination not based on US/BNC rates represent potential limitations.

Conclusions: The midterm safety and efficacy of B-TURP and M-TURP are comparable.

Trial registration: Netherlands Trial Register, NTR703 (http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=703).

Με βάση ποιά κριτήρια γίνεται τυπικά
η αποτίμηση της ποιότητας μιας
Τυχαιοποιημένης Ελεγχόμενης Δοκιμής (RCT);

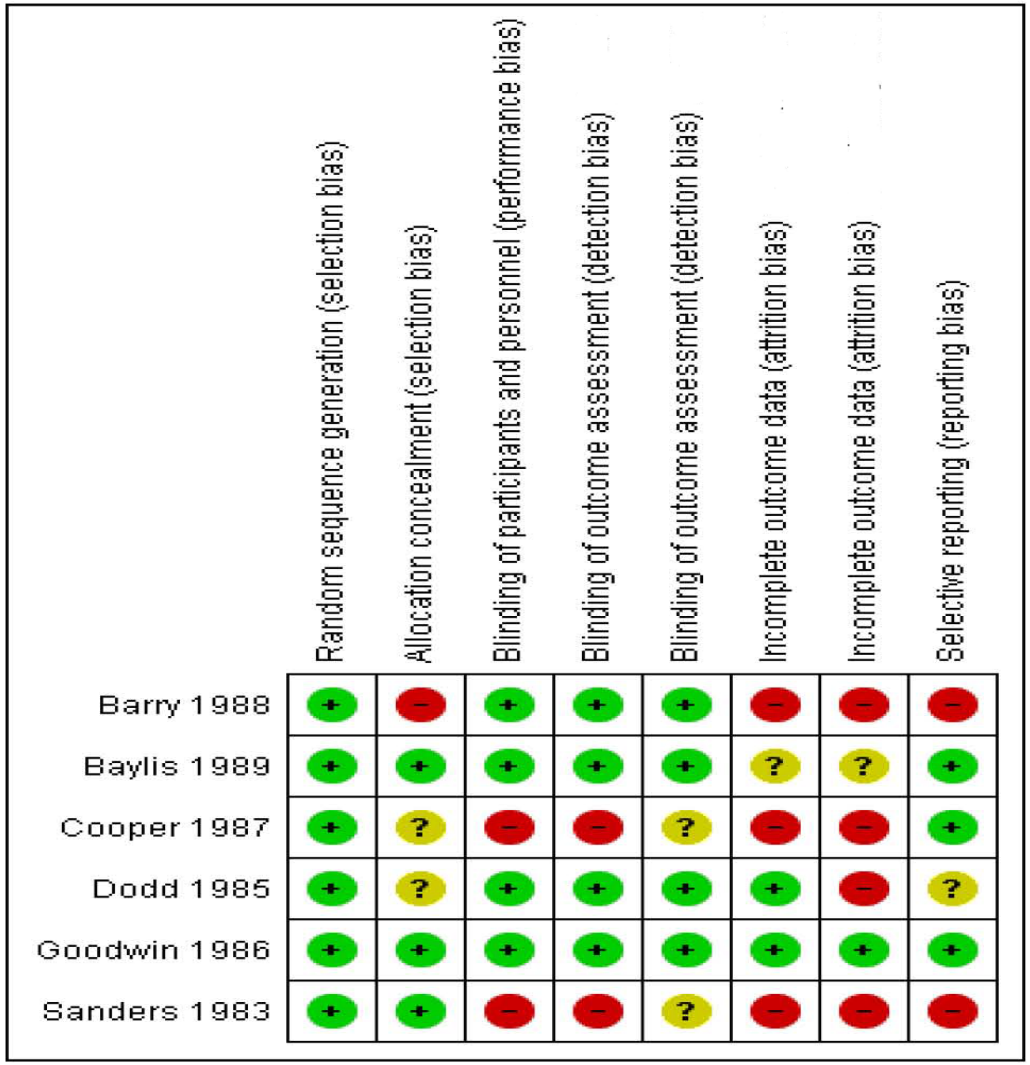
Η αξιοπιστία των αποτελεσμάτων μιας ΤΕΚΔ εξαρτάται από το βαθμό στον οποίο έχουν αποφευχθεί οι δυνητικές πηγές συστηματικού σφάλματος (bias):

- Σφάλμα επιλογής (Selection bias)
- Σφάλμα επίδοσης (Performance bias)
- Σφάλμα ανίχνευσης (Detection bias)
- Σφάλμα ανολοκληρωτης αναφοράς έκβασης (Attrition bias)
- Σφάλμα επιλεκτικής αναφοράς (Selective Reporting bias)

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

8! "#$%&$ '(#) +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr*

Figure 8.6.c: Example of a 'Risk of bias summary' figure



Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Review

BJUI
BJU International

Systematic review and meta-analysis of the clinical effectiveness of bipolar compared with monopolar transurethral resection of the prostate (TURP)

Muhammad Imran Omar, Thomas B. Lam, Cameron E. Alexander, John Graham, Charalampos Mamoulakis*, Mari Imamura, Steven Maclennan, Fiona Stewart and James N'Dow

- Various methodological limitations were highlighted in the included trials and as such the results of this review should be interpreted with caution.

8! "#$%&$ ' (#) * +, ! - ./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Σφάλμα επιλογής (Selection bias) I:

Συστηματικές διαφορές στα βασικά χαρακτηριστικά (baseline characteristics) των προς σύγκριση ομάδων

Αποτρέπεται με την επιτυχή τυχαιοποίηση των συμμετεχόντων στις υπό μελέτη παρεμβάσεις

Η τυχαιοποίηση για να είναι επιτυχής πρέπει να βασίζεται σε κάποιον κανόνα τυχαίας κατανομής των συμμετεχόντων στις ομάδες
(random sequence generation)

RANDOM SEQUENCE GENERATION

Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.

Criteria for a judgement of 'Low risk' of bias.	The investigators describe a random component in the sequence generation process such as: ● Referring to a random number table; ● Using a computer random number generator; ● Coin tossing; ● Shuffling cards or envelopes; ● Throwing dice; ● Drawing of lots;
Criteria for the judgement of 'High risk' of bias.	The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example: ● Sequence generated by odd or even date of birth; ● Sequence generated by some rule based on date (or day) of admission; ● Sequence generated by some rule based on hospital or clinic record number. ● Allocation by judgement of the clinician; ● Allocation by preference of the participant; ● Allocation based on the results of a laboratory test or a series of tests; ● Allocation by availability of the intervention.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'.

Σφάλμα επιλογής (Selection bias) II:

Εξασφάλιση αυστηρής απόκρυψης του κανόνα τυχαιοποίησης στους εμπλεκόμενους στην ένταξη ασθενών (αδύνατη η εκ των προτέρων πρόβλεψη της κατανομής)

Η διαδικασία αυτή λέγεται απόκρυψη (ακολουθίας) κατανομής allocation (sequence) concealment

ALLOCATION CONCEALMENT

Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.

Criteria for a judgement of 'Low risk' of bias.	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: ● Central allocation (including telephone, web-based and pharmacy-controlled randomization); ● Sequentially numbered drug containers of identical appearance; ● Sequentially numbered, opaque, sealed envelopes.
Criteria for the judgement of 'High risk' of bias.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: ● Using an open random allocation schedule (e.g. a list of random numbers); ● Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); ● Date of birth;
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.

A total of 295 patients were enrolled and randomised 1:1 into an M-TURP or B-TURP arm

Randomization: Low Risk

Allocation concealment: Low Risk

Selection bias

Σφάλμα επίδοσης (Performance bias)

Συστηματικές διαφορές μεταξύ των ομάδων στην παρεχόμενη φροντίδα ή την έκθεση σε παράγοντες εκτός των παρεμβάσεων ενδιαφέροντος

Η «Τυφλοποίηση» (blinding) των ασθενών/προσωπικού μειώνει τον κίνδυνο

Η «Τυφλοποίηση» του προσωπικού
δεν είναι πάντα εφικτή
(π.χ. σε χειρουργικές ΤΕΚΔ είναι αδύνατη)

Σφάλμα ανίχνευσης (Detection bias)

Συστηματικές διαφορές μεταξύ των ομάδων στην αξιολόγηση των αποτελεσμάτων

Η «Τυφλοποίηση» των αξιολογητών των αποτελεσμάτων (outcome assessors) μειώνει τον κίνδυνο

BLINDING OF PARTICIPANTS AND PERSONNEL

Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.

BLINDING OF OUTCOME ASSESSMENT

Detection bias due to knowledge of the allocated interventions by outcome assessors.

Criteria for a judgement of 'Low risk' of bias.	Any one of the following: ● No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; ● Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: ● No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; ● Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: ● Insufficient information to permit judgement of 'Low risk' or 'High risk'; ● The study did not address this outcome.

Surgeons were not blinded due to the nature of the interventions. However, outcome assessors were different from the surgeons and the data analyst, who was also not blinded. The outcome assessor at each centre and the patients were blinded for the intervention type.

Blinding of Personnel: High Risk

Blinding of Patients: Low Risk

Performance bias

Blinding of Outcome Assessors: Low Risk

Detection bias

8' "#$%&$ '(#)* +, ! -./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

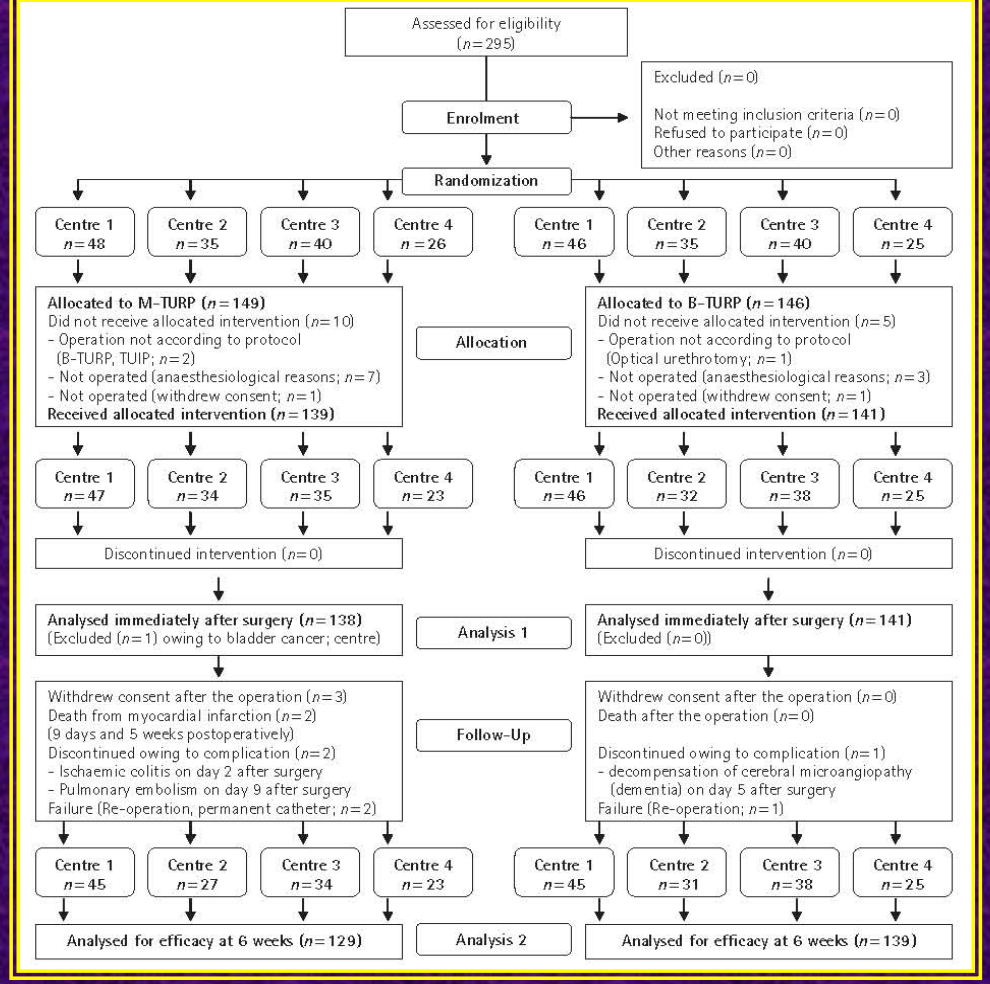
Σφάλμα ανολοκλήρωτης αναφοράς έκβασης (Attrition bias)

Συστηματικές διαφορές μεταξύ των ομάδων στα ποσοστά απόσυρσης ασθενών από τη μελέτη (μη διαθέσιμα δεδομένα – incomplete outcome data)

INCOMPLETE OUTCOME DATA

Attrition bias due to amount, nature or handling of incomplete outcome data.

Criteria for a judgement of 'Low risk' of bias.	● No missing outcome data; ● Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; ● Missing data have been imputed using appropriate methods.
Criteria for the judgement of 'High risk' of bias.	● Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; ● 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; ● Potentially inappropriate application of simple imputation.
judgement of 'Unclear risk' of bias.	● Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided); ● The study did not address this outcome.



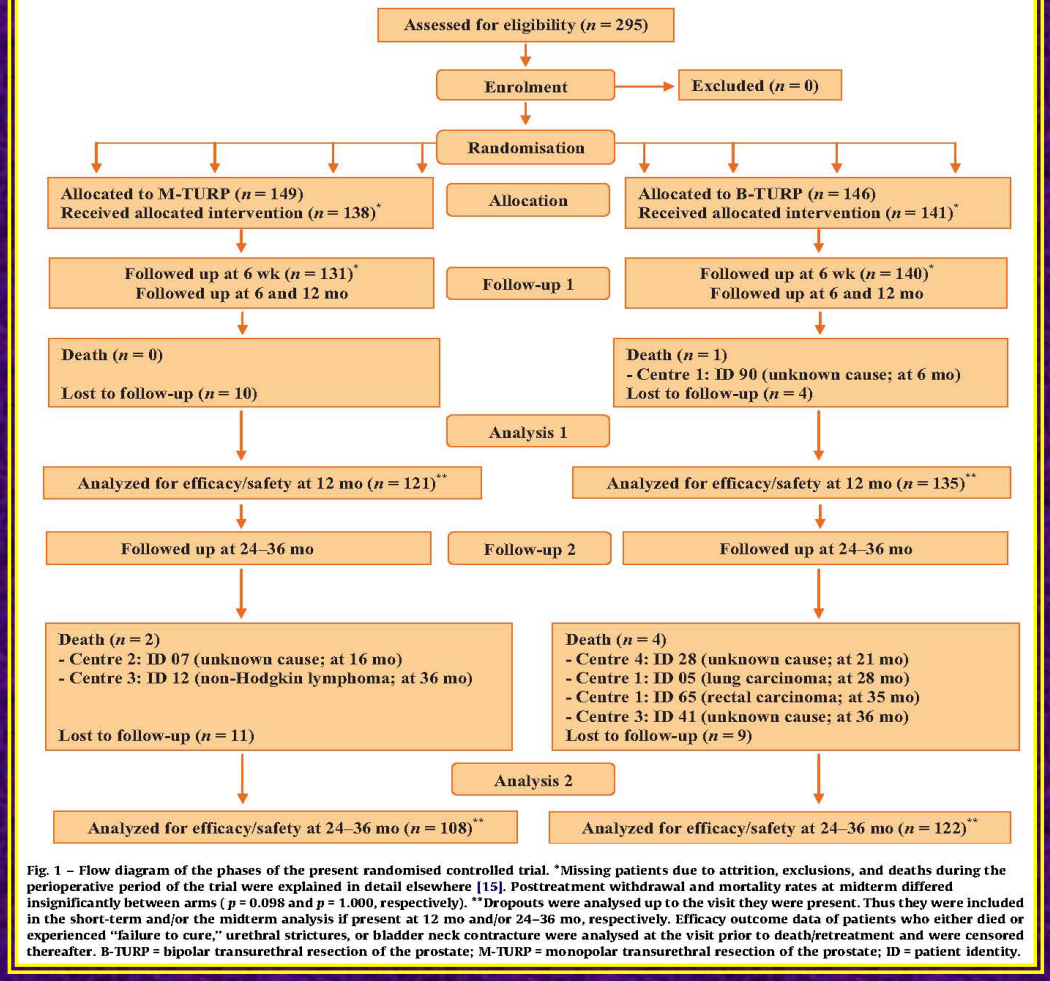


Fig. 1 – Flow diagram of the phases of the present randomised controlled trial. *Missing patients due to attrition, exclusions, and deaths during the perioperative period of the trial were explained in detail elsewhere [15]. Posttreatment withdrawal and mortality rates at midterm differed insignificantly between arms ($p = 0.098$ and $p = 1.000$, respectively). **Dropouts were analysed up to the visit they were present. Thus they were included in the short-term and/or the midterm analysis if present at 12 mo and/or 24–36 mo, respectively. Efficacy outcome data of patients who either died or experienced "failure to cure," urethral strictures, or bladder neck contracture were analysed at the visit prior to death/retreatment and were censored thereafter. B-TURP = bipolar transurethral resection of the prostate; M-TURP = monopolar transurethral resection of the prostate; ID = patient identity.

Incomplete Outcome Data: Low Risk → Attrition bias

8' "#$%&$ ' (#) * +, ! - ./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Σφάλμα επιλεκτικής αναφοράς (Selective Reporting bias)

Συστηματικές διαφορές μεταξύ ανακοινωθέντων και μη ανακοινωθέντων ευρημάτων

Συνηθισμένο φαινόμενο η συχνότερη ανακοίνωση των στατιστικά σημαντικών έναντι των μη στατιστικά σημαντικών αποτελεσμάτων («within-study publication bias»)

Αποτελεί ένα από τα πιο ουσιαστικά σφάλματα

SELECTIVE REPORTING

Reporting bias due to selective outcome reporting.

Criteria for a judgement of 'Low risk' of bias.	Any of the following: ● The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; ● The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'High risk' of bias.	Any one of the following: ● Not all of the study's pre-specified primary outcomes have been reported; ● One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; ● One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); ● One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; ● The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.

A prospective, randomized, double-blinded study to compare bipolar Trans Urethral Resection of the Prostate (bipolar TURP) versus monopolar Trans Urethral Resection of the Prostate (monopolar TURP) in terms of safety and efficacy.

- TRIAL ID NTR703

NEDERLANDS TRIAL REGISTER

Bipolar vs monopolar transurethral resection of the prostate: evaluation of the impact on overall sexual function in an international randomized controlled trial setting

Charalampos Mamoulakis[1,2], Andreas Skolarikos[3], Michael Schulze[4], Cesare M. Scoffone[5], Jens J. Rassweiler[4], Gerasimos Alivizatos[3], Roberto M. Scarpa[5] and Jean J.M.C.H. de la Rosette[1]

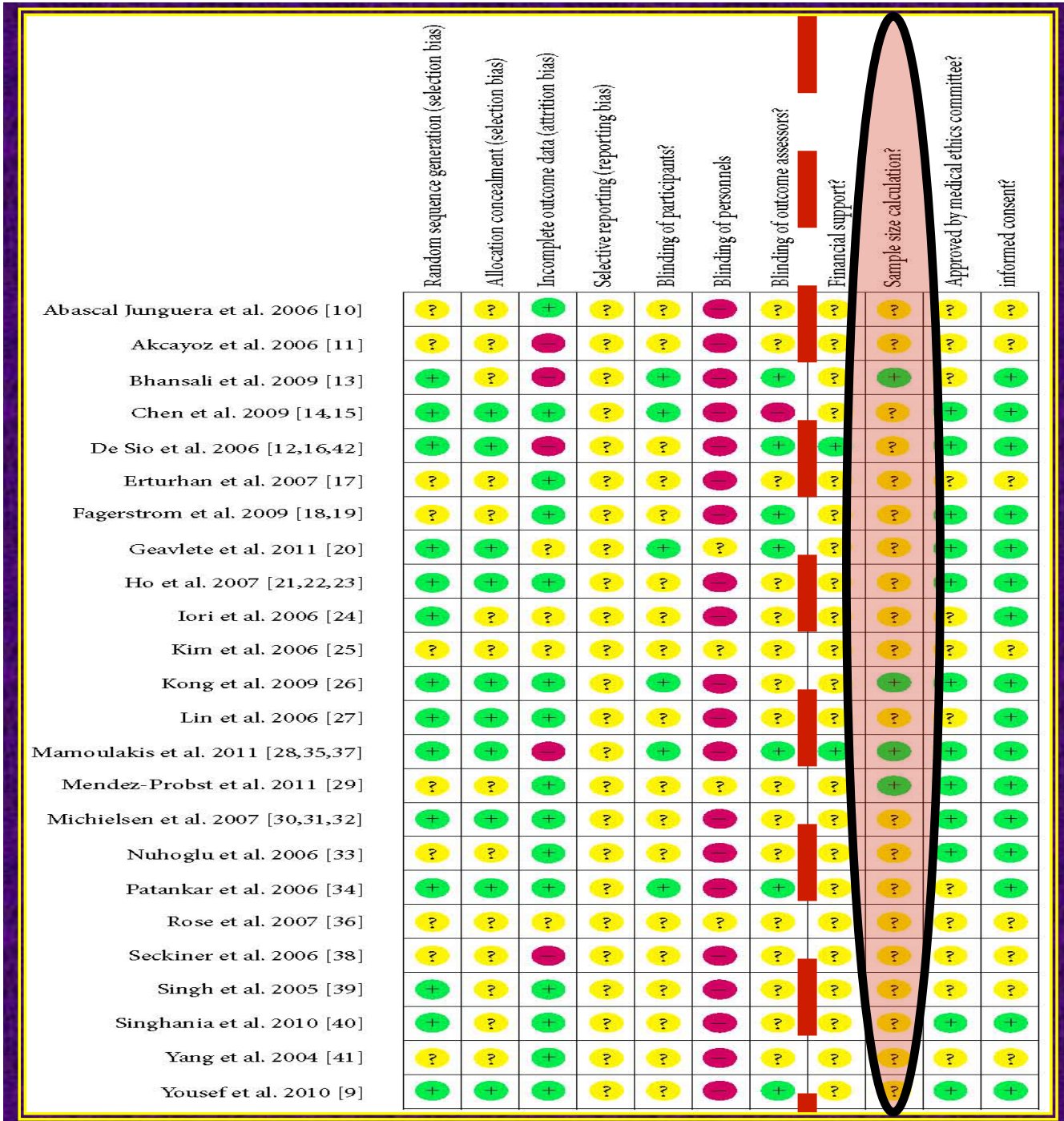
Mamoulakis et al. BJU Int. 2013;112:109-20

Patients were evaluated (IIEF-15): baseline, 6wk, 6,12, 24, and 36 mo

The impact on overall sexual function does not differ significantly between M-TURP & B-TURP at 12 mo

Selective Reporting: High Risk → Reporting bias

8ι "#$%&$ '(♯)* +, ! -./! 01&! - !2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr



Άλλες πηγές Σφάλματος:

Υπολογισμός Μεγέθους Δείγματος

Τυχαία Σφάλματα

8ˡ "#$%&$ '(#)* +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr

Εξαρτόνται από τη διακύμανση των δεδομένων και το μέγεθος δείγματος

Μια μέτρηση vs. πολλές μετρήσεις ενός μεγέθους: ακριβέστερη εκτίμηση της πραγματικότητας

Ο βαθμός αβεβαιότητας/τυχαίου σφάλματος εκφράζεται με τα διαστήματα εμπιστοσύνης (Confidence Interval)

Εαν τα δεδομένα μετρηθούν άπειρες φορές το συμβατικό 95% CI θα περιλαμβάνει την πραγματική τιμή
95% των φορών
Στενό διάστημα εμπιστοσύνης = μεγαλύτερη ακρίβεια
(επιτυγχάνεται με μεγαλύτερο δείγμα)

Τύπου I

Βλέπουμε μια στατιστικά σημαντική διαφορά ενώ δεν υπάρχει στην πραγματικότητα
(απόρριψη Ho ενώ ισχύει)

Τύπου II

Δεν βλέπουμε μια στατιστικά σημαντική διαφορά ενώ υπάρχει στην πραγματικότητα
(αποδοχή Ho ενώ δεν ισχύει)
(Συχνά σφάλματα: Ανεπαρκές μέγεθος δείγματος)

Υπολογισμός Μεγέθους Δείγματος

Σημασία:

- Λογική βεβαιότητα ανίχνευσης σημαντικότητας αν υπάρχει (αποφυγή σφάλματος τύπου II): Ισχύς (Power)
- Μείωση δαπάνης εφεδριών (χρόνος – χρήματα)
- Θέματα ηθικής

Υπολογισμός Μεγέθους Δείγματος

Τί απαιτείται για να υπολογιστεί:

- Θέτω την επιθυμητή πιθανότητα σφάλματος τύπου I (α) (επίπεδο στατιστικής σημαντικότητας ενδεικνυόμενου test P value: α = 0,05/ α = 0,01/ α = 0,001)
- Θέτω την επιθυμητή πιθανότητα σφάλματος τύπου II (β)
- (επίθυμητή ισχύς μελέτης: power (1-β) ≥ 0,8
- Θέτω τη διαφορά που ενδιαφέρει να ανιχνευθεί αν υπάρχει

Table 1

Factors that affect sample size calculations

Factor	Magnitude	Impact on identification of effect	Required sample size
P value	Small	Stringent criterion; difficult to achieve 'significance'	Large
	Large	Relaxed criterion; 'significance' easier to attain	Small
Power	Low	Identification unlikely	Small
	High	Identification more probable	Large
Effect	Small	Difficult to identify	Large
	Large	Easy to identify	Small

Υπολογισμός Μεγέθους Δείγματος

Τρόπος υπολογισμού:

- Τί ειδους μεταβλητή μας ενδιαφέρει;
  (primary study outcome): κατηγορική, συνεχής, χρονική
- Τί είδους ομάδες έχουμε;
  (1 ομάδα: πριν και μετά την παρέμβαση - 2 ανεξάρτητες)
- Επιλογή κατάλληλης δοκιμασίας με βάση τα παραπάνω
- Είναι οι ομάδες ίσες; (1:1) ή άνισες (2:1 κ.λ.π.)
- Θα υπολογίσουμε extra απώλειες (drop out rate?)
- Χρήση αντίστοιχων τύπων/νομογραμμάτων

Τί ειδους στατιστική δοκιμασία θα χρησιμοποιούσατε;

Για να αναλύσετε:

- Φύλο (σε 2 ομάδες)

  Κατηγορική μεταβλητή: Pearson X2 – Fisher's X2

- Προεγχειρητικό Qmax (σε 2 ομάδες / σε 3 ομάδες)

  Συνεχής μεταβλητή

  T-test ή Mann Whitney (??) / ANOVA ή Kruskal Wallis

- ΔQmax (Προ-Μετεγχειρητικό σε 2 ομάδες / σε 3 ομάδες)

  Συνεχής μεταβλητή

  T-test paired ή Wilcoxon (??) / Repeated measure ANOVA

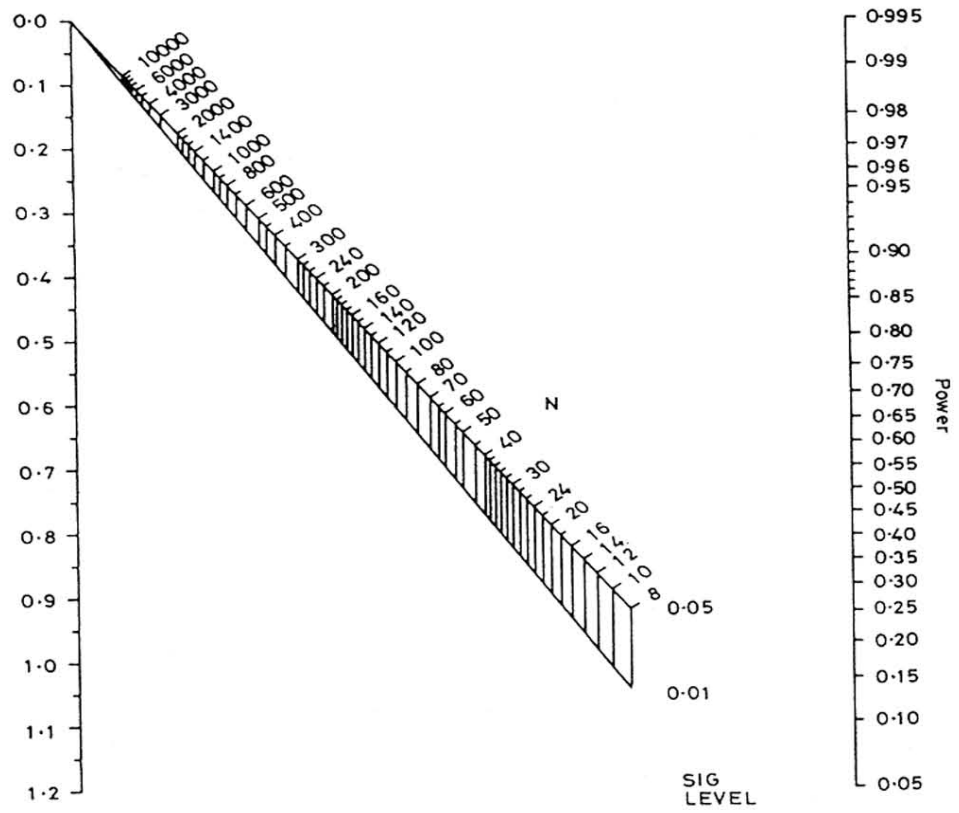
- QoL score (σε 2 ομάδες / σε 3 ομάδες)

  Ταξινομική μεταβλητή

  Mann Whitney ή Kruskal Wallis

Υπολογισμός Μεγέθους Δείγματος (πρακτικά θέματα)

Figure 1

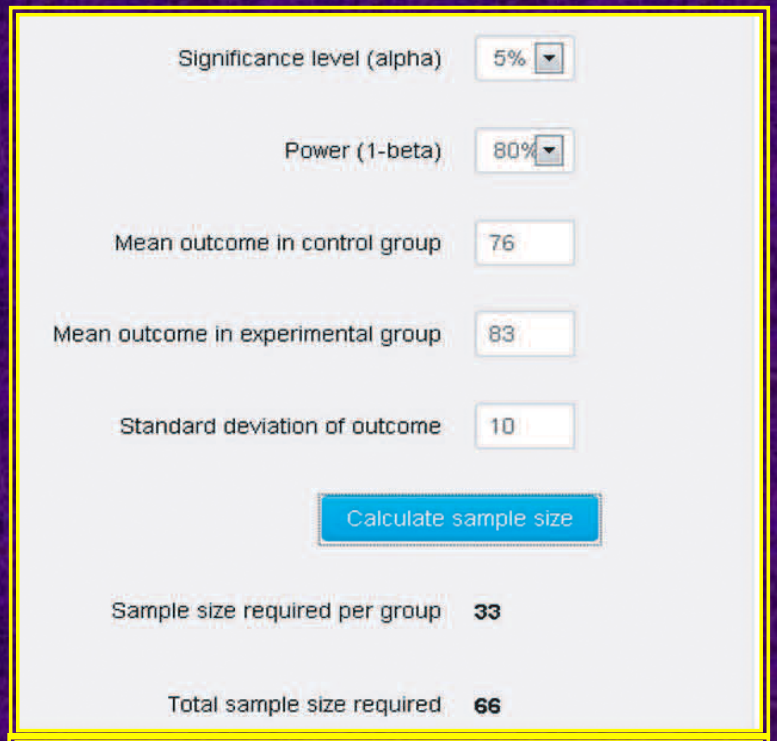


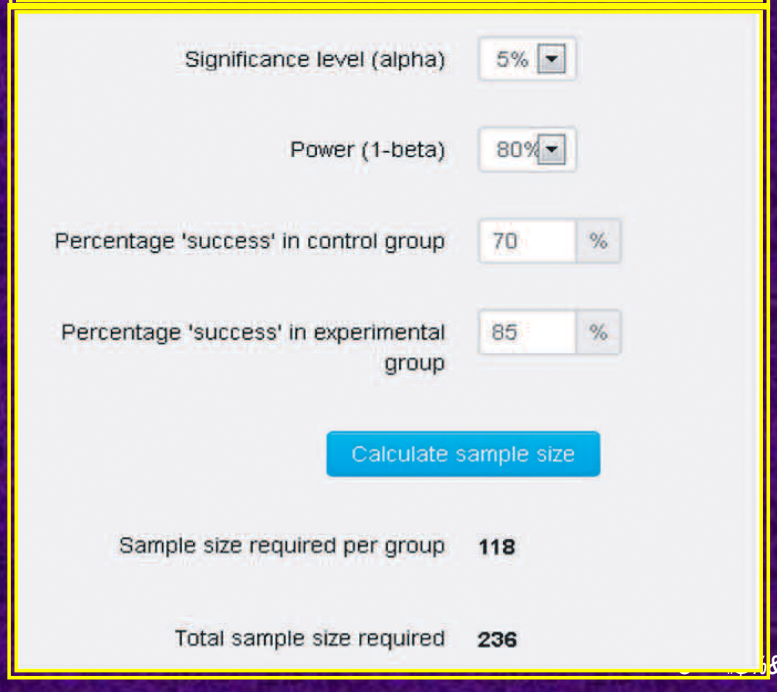
$$\text{Standardized difference} = \frac{\text{Target difference}}{\text{Standard deviation}}$$

$$\text{Standardized difference} = \frac{(p_1 - p_2)}{\sqrt{[\bar{p}(1 - \bar{p})]}} \qquad (3)$$

where p_1 and p_2 are the proportions in the two groups and $\bar{p} = (p_1 + p_2)/2$ is the mean of the two values.

$$N' = \frac{N(1 + k)^2}{4k}$$





Programs

- G*Power: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/
  G*Power is a representative program for the sample size calculation, but it is hard to use and the instruction manual is also difficult to find. You can refer to "*Sample size calculation* (ISBN 9788994467764)".
- Piface: http://homepage.stat.uiowa.edu/~rlenth/Power/ (free)
- PS: http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize (free)
- NQuery Advisor: Program for calculating the number of samples (charged)
- R (free), Medcalc (charged), SAS (charged), Expansion of SPSS (charged): Statistical program which can calculate the sample size also.

Websites

- http://department.obg.cuhk.edu.hk
- http://www.quantitativeskills.com/sisa/calculations/samsize.htm
- http://www.statstodo.com/SSizSurvival_Pgm.php
  The links listed above have a function of common statistical solutions as well as sample size calculation.

- http://www.cct.cuhk.edu.hk/stat/Proportions.htm
- http://www.sealedenvelope.com/power/
- http://www.dartmouth.edu/~eugened/power-samplesize.php
- http://cafe.naver.com/easy2know/6259
- http://www.danielsoper.com/statcalc3/calc.aspx?id=5
  The links listed above are only for calculating the sample size.

&$ '(#) +, ! -./! 01&! -!2/$3, 5-8 4$&(/!1 2015, 5. 4$µ!1-6)73: c.mamoulakis@med.uoc.gr*

Εσφαλμένη χρήση στατιστικών δοκιμασιών στην Ουρολογία

0022-5347/05/1744-1374/0
THE JOURNAL OF UROLOGY®

Vol. 174, 1374–1379, October 2005
Printed in U.S.A.
DOI: 10.1097/01.ju.0000173640.91654.b5

CLINICAL RESEARCH AND STATISTICAL METHODS IN THE UROLOGY LITERATURE

CHARLES D. SCALES, JR., REGINA D. NORRIS, BERCEDIS L. PETERSON, GLENN M. PREMINGER AND PHILIPP DAHM*

ABSTRACT

Purpose: We provide a systematic assessment of the quality and accuracy of statistical reporting in the urology literature.

Materials and Methods: All original research publications with adult human subjects in a single issue (August 2004) of 4 leading urology journals were identified for formal review. A standardized evaluation form was developed in consultation with an experienced biostatistician and subsequently tested. Two independent reviewers with at least 1 year of formal training in research design and biostatistics who were blinded to authors and institutions reviewed each article. Discrepancies were settled by consensus and/or adjudication by the biostatistician.

Results: Of the 169 articles screened 97 met eligibility criteria for review. Cohort (43 of 97 or 44%) or cross-sectional (28 of 97 or 29%) designs comprised the majority of these studies. Only 10 randomized clinical trials (12.4%) were identified. Statistical tests were identified in 83 studies (93%). Overall 69 of 83 studies (71%) providing statistical comparisons had at least 1 statistical error, including using the wrong test for the data type in 28%, inappropriate use of a parametric test in 22% and failure to account for multiple comparisons in 65%. In studies applying multivariate analysis (29%) over fitting the model with too many variables was the most common statistical flaw (39%).

Εσφαλμένη χρήση στατιστικών δοκιμασιών στην Ουρολογία

TABLE 5. *Specific statistical errors identified on systematic review*

Biostatistical Error Type	No. With Error	No. Applicable Studies (%)
Any statistical error*	69	83 (71)
Inappropriate data test (any)*	22	80 (28)
Categorical (continuous error)	6	80 (8)
Paired (unpaired error)	6	71 (9)
Parametric (nonparametric error)	17	77 (22)
Pairwise comparison without prior testing for difference among all groups	9	27 (33)
Failure to account for multiple testing	47	72 (65)
Regression errors:		
All	23	28 (82)
Model over fitting	11	28 (39)

Conclusions: This formal review suggests that statistical methods are often used inappropriately in the urology literature, thereby, potentially undermining the validity of study results and conclusions. An effort to raise the awareness of appropriate statistical techniques through postgraduate education appears indicated.


ΕΥΧΑΡΙΣΤΩ