DJOG Danish Journal of Obstetrics and Gynaecology

Baseline measurements can be omitted in urethral and anal pressure crossover studies

Christoffersen T.[1,2*], Kornholt J. [1*], Riis T.[1], P. Sonne D. P.[1,2], Klarskov N.[2]

[1]Department of Clinical Pharmacology, Copenhagen University Hospital - Bispebjerg and Frederiksberg, Bispebjerg Bakke 23, 2nd floor, 2400 Copenhagen, Denmark
[2]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[3]Department of Gynecology and Obstetrics, Copenhagen University Hospital - Herlev and Gentofte, Borgmester Ib Juuls Vej 1, 2730 Herlev, Denmark
**Currently employed at Novo Nordisk A/S, Bagsvaerd, Denmark*

Corresponding author: Niels Klarskov; niels.klarskov@regionh.dk

ABSTRACT

Objectives: To explore the impact of baseline measurements on variance, the precision of treatment estimate, and sample size in crossover studies using the urethral pressure and anal acoustic reflectometry methodologies.

Methods: This was a post hoc analysis of a randomized, double-blind, placebo-controlled, crossover study of the effect of imipramine on urethral and anal opening pressure. We applied three analysis-of-covariance models that include baseline measurements in the three most common ways and performed sample size calculations for future crossover studies based on the within-subject variance from the three models.

Results: The model which ignores the baseline measurement provided the lowest variance and thus the highest precision of treatment estimate and the smallest sample size whereas the model that incorporates baseline measurements as a change from baseline analysis provided the largest variance, lowest precision of treatment estimate, and largest sample size estimation.

Conclusion: Our findings suggest that it is not beneficial to include baseline measurements in crossover studies with urethral pressure and anal acoustic reflectometry.

Keywords: Randomised controlled trial; Crossover studies; Urethral pressure reflectometry; Anal acoustic reflectometry; Urethra; Anal canal

Received: 18. April 2024
Accepted: 10. June 2025
Date of publication: 11.June 2025
DOI: https://doi.org/10.56182/vtyhyj89

INTRODUCTION

Crossover studies are used in various disciplines within medical research, particularly for the evaluation of pharmacological treatments for chronic, stable conditions, and in pharmaceutical phase 1 studies. The benefit of the crossover design is that the study subjects serve as their own controls, thereby eliminating between-subject variance and reducing the required sample size. Baseline measurements of the outcome variable may be included in crossover studies, either as a single measurement at the beginning of the study (pre-randomization baseline) or measurements obtained before treatment in each period (within-period baseline). However, the advantage of including baseline measurements in crossover studies, as well as the methodology of incorporating baseline measurements into the analyses has been widely discussed in the literature (1-5). Three common approaches include i) using a change from baseline analysis, ii) incorporating baseline values as a covariate in an analysis of covariance (ANCOVA), and iii) ignoring baseline measurements. In general, most statisticians recommend the inclusion of within-period baseline measurements if the crossover study is unbalanced and/or missing data are expected or if there is reason to believe that the baseline measurements contain additional information not accounted for by including subject and period effects (1, 2, 5, 6).

Urethral pressure reflectometry (UPR) is a measurement technique that uses an ultra-thin inflatable polyurethane bag placed in the urethra to measure simultaneously the cross-sectional area and urethral pressure by acoustic reflectometry (7). Previously, it has been demonstrated that UPR performs better in terms of sensitivity and reproducibility compared with conventional urethral pressure profilometry (8, 9). The technique has later been adopted for use in the anal canal (anal acoustic reflectometry [AAR]) by Mitchell et al. (10, 11). Recently, opening urethral pressure (OUP) measured with UPR and anal opening pressure (AOP) measured with AAR have been introduced as primary outcome measures in single-dose pharmacodynamic crossover studies (9, 12, 13). In the first studies, it was demonstrated that UPR is capable of detecting drug-induced pressure changes in the urethra and that the pressure increases assessed in these studies appear to correlate well with the clinical effect of the drugs tested (9, 14).

In these first studies, repeated UPR measurement sessions were performed in each treatment period, including a predose measurement session, to investigate how changes in plasma concentration coincided with changes in urethral pressure. These studies revealed that the maximal increase in urethral pressure was very close to the time point nearest the time of maximal plasma concentration (T_{max}) and that no placebo effect was observed (12). Consequently, in subsequent crossover studies the outcome measurement was reduced to a single measurement session corresponding to the estimated T_{max} of the study drug (15, 16). However, the impact of including predose measurements as within-period baselines on within-subject variance in UPR and AAR crossover studies has not yet been assessed. It is possible that individual trend effects, such as pressure changes due to factors other than the study drugs, may have affected participants' urethral and/or anal pressure during the study period. In this case, baseline data might reduce the within-subject variance, which would increase the precision of the treatment estimate. Therefore, we aimed to assess the impact of baseline measurements by conducting a post hoc analysis of data from our randomized, placebo-controlled crossover study assessing the effect of imipramine on urethral and anal pressure. Our overall purpose is to provide recommendations regarding design and analysis of future UPR and AAR crossover studies.

MATERIALS AND METHODS

We evaluated the impact of baseline measurements on the within-subject variance, the precision of the treatment estimate, and on sample size calculation by a post hoc analysis of data from our randomized, double-blind, placebo-controlled crossover study studying the effect of imipramine on urethral and anal pressure in healthy female volunteers. The results from the urethral assess-

ISSN 2794-3372
2025, Volume 3, page 97

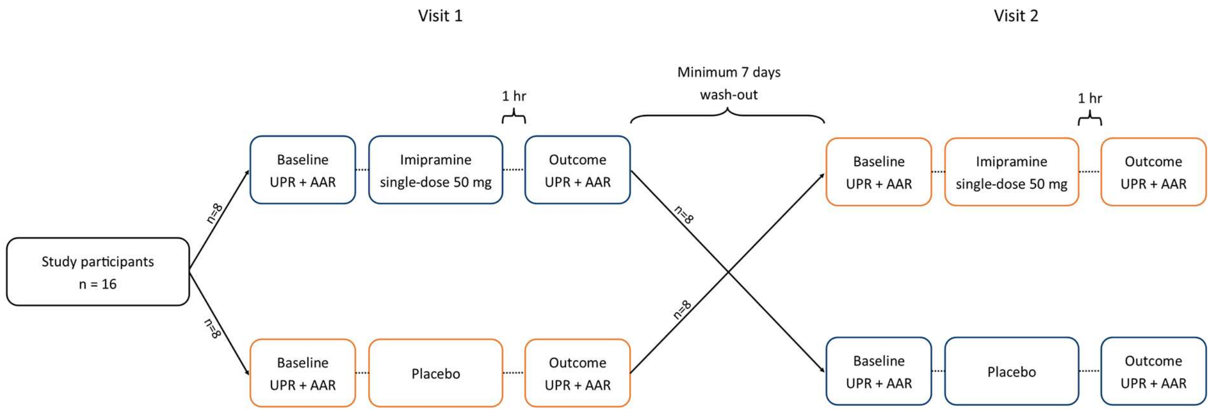

Figure 1: Study design for the randomized, double-blind, placebo-controlled crossover study assessing the effect of imipramine on urethral and anal pressure in healthy female volunteers. UPR, urethral pressure reflectometry; AAR, anal acoustic reflectometry

ments and a detailed description of the study design and methods were previously published (15). In short, 16 healthy women were recruited, screened and randomized (1:1) to one of two sequences. In sequence A, the participants (n=8) received a single dose of 50 mg imipramine at their first visit and a matched placebo at their second visit, and vice versa in sequence B (n=8). At both visits, UPR and AAR measurements were performed at baseline (within-period baseline) and one hour after the administration of imipramine/placebo (corresponding to the estimated T_{max} of imipramine) (Figure 1).

Table 1 Specification of variables in the three ANCOVA models

Models	Dependent variable	Explanatory variables (covariates)
Model 1 - Ignoring baseline	1 hr OUP/AOP	Subject Period (day) Treatment
Model 2 - Baseline adjusted	1 hr OUP/AOP	Subject Period (day) Treatment Baseline OUP/AOP
Model 3 - Change from baseline	Difference between baseline and 1 hr OUP/AOP	Subject Period (day) Treatment

OUP, opening urethral pressure; AOP, anal opening pressure.

Statistical methods

To assess the impact of baseline measurements on the within-subject variance, precision of treatment estimate and thereby sample size, we applied three analyses of covariance (ANCOVA) models, which incorporate the baseline measurements in different ways (Table 1):

Model 1: Ignoring baseline

In this model, baseline measurements are omitted from the analysis and the one-hour measurement is used as dependent variable.

Model 2: Baseline adjusted

In this model, the one-hour measurement is used as dependent variable, and the baseline measurements are fitted as covariate in the ANCOVA model.

Model 3: Change from baseline

This model uses the difference between baseline and the one-hour measurement at the same visit as dependent variable to estimate the treatment effect of imipramine vs. placebo.
For all analyses, we used the proc glm procedure in SAS® with subject, treatment, and period fitted as fixed effects (1,6). We did not use mixed models as we did not have missing data. Further, we assumed that no carry-over effect had biased the measurements because a long wash-out period

ISSN 2794-3372
2025, Volume 3, page 98

Table 2: Output from the three ANCOVA models

Outcome	Model	Treatment estimate (imipramine vs placebo, cmH_2O)	95% Confidence Interval	Standard Error[a] (SE)	Sample size calculation[c]		
					RMSE[b]	σ	n
OUP Resting condition	Model 1 - Ignoring baseline	5.59	(2.13; 9.06)	1.61	4.56	6.45	6
	Model 2 - Baseline adjusted	5.44	(1.92; 8.96)	1.63	4.58	6.48	6
	Model 3 - Change from baseline	6.53	(-0.48; 13.53)	3.27	9.24	13.07	18
OUP Squeezing condition	Model 1 - Ignoring baseline	7.16	(2.32; 12.01)	2.26	6.39	9.04	10
	Model 2 - Baseline adjusted	7.24	(2.23; 12.25)	2.32	6.55	9.26	10
	Model 3 - Change from baseline	7.94	(-0.25; 16.14)	3.82	10.81	15.29	26
AOP Resting condition	Model 1 - Ignoring baseline	9.51	(-2.86; 21.87)	5.77	16.31	23.07	24
	Model 2 - Baseline adjusted	11.54	(-1.92; 25.0)	6.23	16.42	23.22	26
	Model 3 - Change from baseline	15.12	(2.01; 28.21)	6.11	17.27	24.42	28
AOP Squeezing condition	Model 1 - Ignoring baseline	11.24	(3.69; 18.78)	3.52	9.95	14.07	10
	Model 2 - Baseline adjusted	11.28	(3.08; 19.47)	3.79	10.32	14.59	12
	Model 3 - Change from baseline	15.11	(4.25; 25.96)	5.06	14.31	20.24	20

[a], Standard Error of treatment estimate; b, RMSE denotes root mean square error from the ANCOVA models, which corresponds to the within-subject standard error; c, Sample size calculation based on RMSE as variance estimate, a clinically relevant difference of 10 cmH_2O for OUP and 15 cmH_2O for AOP, an alpha level of 0.05, and beta level of 0.2.
OUP rest, opening urethral pressure under resting condition of the pelvic floor; OUP squeeze, opening urethral pressure measured under squeezing condition of the pelvic floor; AOP rest, anal opening pressure measured under resting condition of the pelvic floor; AOP squeeze, anal opening pressure measured under squeezing condition of the pelvic floor.

between the two study days (a minimum of seven days, ~9 half-lives of imipramine) was ensured.
Finally, we estimated the sample size for future two-treatment, two-period (2 x 2) crossover studies using the within-subject variance derived from each of the ANCOVA models (model 1-3). For these sample size calculations, we used the SAS® code provided by Senn (1).
For sample size calculations, we defined the minimal clinically relevant difference for UPR as 10 cmH_2O, based on the pivotal pharmacodynamic study by Klarskov et al. (12), which reported placebo-corrected mean increases in opening urethral pressure (OUP) of 9–46 cmH_2O following administration of midodrine, duloxetine, and reboxetine. The smallest increase (9 cmH_2O with midodrine) has been associated with minor clinical improvement in stress incontinence (17), while larger effects align with duloxetine's proven efficacy in randomized trials (18). For anal opening pressure (AOP), no prior pharmacodynamic studies exist. We therefore used a target difference of 15 cmH_2O, informed by a previous study showing a mean AOP difference of ~20 cmH_2O (SD 26) between women with and without fecal incontinence (19), acknowledging the lack of pharmacological intervention in that study.
All statistical analyses for this paper were generated using SAS software, version 7.15 of the SAS system for Windows (© Copyright 2017, SAS Institute Inc., Cary, North Carolina, USA) and graphical plots using GraphPad Prism version 9.4.1 for Windows (GraphPad Software, San Diego, California, USA, www.graphpad.com).

Ethical approval

The study was approved by the Ethics Committee of the Capital Region of Denmark (approval number H-17007330) and by the Danish Medicines Agency (EudraCT number 2017-000119-18) and was registered at ClinicalTrials.gov (identifier NCT03102645).

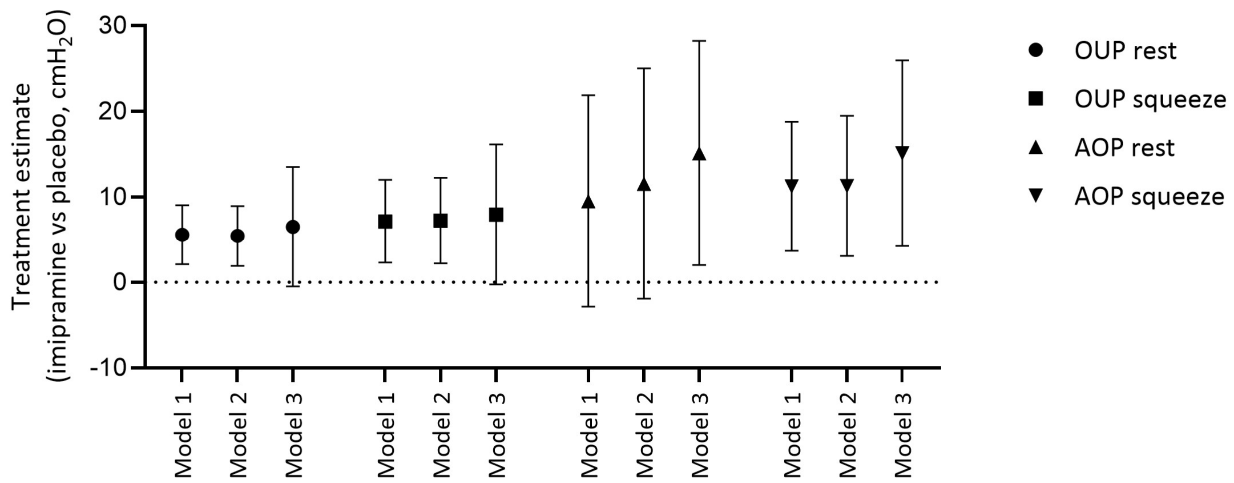

Figure 2: Treatment estimates with 95% confidence intervals from the three ANCOVA models. In Model 1 (Ignoring baseline), baseline values are omitted, and the one-hour measurement is used as dependent variable. In Model 2 (Baseline adjusted), the one-hour measurement is used as dependent variable, and the baseline measurements are fitted as a covariate in the ANCOVA model. In Model 3 (change from baseline), the difference between baseline and the one-hour measurement at the same visit is used as dependent variable.
OUP rest, opening urethral pressure measured under resting condition of the pelvic floor; OUP squeeze, opening urethral pressure measured under squeezing condition of the pelvic floor; AOP rest, anal opening pressure measured under resting condition of the pelvic floor; AOP squeeze, anal opening pressure measured under squeezing condition of the pelvic floor.

RESULTS

The baseline characteristics of the study participants and other study-related information can be found in the original paper (15).
With regards to the variance of the treatment estimate (standard error of the treatment estimate [SE] and the within-subject standard error [RMSE]) model 3 (change from baseline) generated the highest variance of the three models both for urethral and anal measurements and across measurement conditions (resting and squeezing). There was only a minor change in treatment estimates, 95% confidence intervals, and variance between model 2 (baseline adjusted) and model 1 (ignoring baseline) (Table 2 and figure 2). Sample sizes for future two-treatment, two-period (2x2) crossover studies calculated based on the within-subject variance from the three analysis models are presented in the right-hand side of Table 2. For studies using only UPR measurements as outcome, the estimated sample size for model 1 or 2 is more than halved compared with model 3 (change from baseline). Model 2 (baseline adjusted) did not influence sample size compared to a model that omits baseline data (model 1). A similar pattern was seen for AAR data with a 50% reduction in sample size for AOP squeeze. However, for AOP resting data the reduction in sample size from model 3 (change from baseline) to model 1 (ignoring baseline) was only 14% (from n=28 to n=24).

DISCUSSION

In this post hoc analysis of a pharmacodynamic crossover study using UPR and AAR measurements as primary outcome, we evaluated the impact of baseline measurements on variance, precision of treatment estimate and sample size. The outputs from the ANCOVA model 1-3 showed that the model which ignores baseline measurements (model 1) provided the smallest variance and sample size, whereas model 3, in which the change from baseline was applied as the dependent variable, provided the largest variance. These findings suggest that the variance is not reduced by incorporating baseline measurements in the analysis of these types of crossover studies – neither in a change-from-baseline analysis nor in a baseline-as-covariate ANCOVA model. Furthermore, our analyses indicate that the sample size based on model 3 is more than two-fold higher compared with a sample size calculation on model 1 or 2.

ISSN 2794-3372
2025, Volume 3, page 100

Our findings are supported by the general considerations regarding baseline measurements in two key texts on the design and analysis of clinical crossover studies (1, 6). Senn argues that baseline measurements should only be incorporated in the outcome analysis of simple, two-treatment, two-period crossover studies if it is expected that the baseline measurements contain valuable information not already accounted for (by adjustment for subject and period effect) and if carry-over effects can be considered eliminated once the baseline measurement in the second period has been obtained (1). Similarly, Jones and Kenward argue that baseline measurements generally do not increase the precision of the treatment estimates in balanced, complete crossover designs (6). The design in the present study (as well as all previous UPR pharmacodynamic crossover studies) is a simple balanced design with complete data and no expected carry-over effect due to a single-dose design and adequate wash-out. With the abovementioned design characteristics, this study does not fulfil the criteria for crossover design that, in theory, would benefit from the incorporation of baseline measurements. When discussing the inclusion or exclusion of baseline measurements in this type of study, ethical aspects must also be considered. Although UPR and AAR are generally well-tolerated, they are invasive procedures and carry potential risks and adverse events, including discomfort, pain, and a low probability of urinary tract infections (related to urethral measurements). Therefore, it would be beneficial to minimize the number of measurement sessions during the study period. In a two-period study, within-period baseline measurements constitute two additional measurements, and it appears most ethical to exclude them.

Our post hoc analysis has some limitations. The imipramine study was not designed to evaluate the importance of baseline measurements in crossover studies. This increases the risk of random or spurious findings. Furthermore, we limited the handling of baseline measurements in crossover studies to three commonly used approaches. Several other approaches have been proposed. For example, Kenward and Roger suggested incorporating baseline measurements as response variables without associated treatment effects in a mixed model with random subject effects (2). However, they found no difference in estimates and variance compared with the baseline-as-covariate approach in simple, balanced studies using a model with fixed subject effect. Consequently, we believe it is unlikely that such a model would perform better with our data. Overall, there is no consensus on the most effective way to incorporate baseline measurements in crossover studies. Therefore, we have utilized the most common approaches and available data to evaluate the impact of baseline measurements in UPR and AAR crossover studies. Our findings indicate that including baseline measurements does not appear to be advantageous in these types of studies.

CONCLUSION

The objective of this paper was to examine how baseline measurements affect the variance, precision of treatment estimates, and sample size in crossover studies utilizing the UPR and AAR methodologies. Considering theoretical arguments, empirical results from our analysis that demonstrate a lower variance resulting in increased precision and reduced sample size when excluding baseline measurements, and ethical considerations, we recommend that future UPR and AAR crossover studies exclude baseline measurements from their design.

Conflict of interest: The authors declare that they have no conflicts of interest. Thea Christoffersen and Jonatan Kornholt are employees of Novo Nordisk, Denmark. However, the ideas and opinions discussed in this publication are those of the authors and might not reflect those of their employer.
Funding information: No specific funding was obtained for this research.
Acknowledgements: None.

REFERENCES

1. Senn S. Cross-over trials in clinical research. 2nd ed. Chichester; JWiley; 2002. 363 p.

2. Kenward MG, Roger JH. The use of baseline covariates in crossover studies. Biostatistics. 2010;11(1):1-17.

3. Mehrotra DV. A recommended analysis for 2 × 2 crossover trials with baseline measurements. Pharmaceut Statist. 2014;13(6):376-87.

4. Xu R, Mehrotra DV, Shaw PA. Incorporating baseline measurements into the analysis of cross-over trials with time-to-event endpoints. Statistics in Medicine. 2018;37(23):3280-92.

5. Yan Z. The impact of baseline covariates on the efficiency of statistical analyses of crossover designs. Statistics in Medicine. 2013;32(6):956-63.

6. Jones B, Kenward MG. Design and analysis of cross-over trials. Third edition ed. Boca Raton: CRC Press/Taylor & Francis; 2014. 412 p.

7. Klarskov N, Lose G. Urethral pressure reflectometry; a novel technique for simultaneous recording of pressure and cross-sectional area in the female urethra. Neurourology and Urodynamics. 2007;26(2):254-61.

8. Klarskov N, Lose G. Urethral pressure reflectometry vs urethral pressure profilometry in women: a comparative study of reproducibility and accuracy. BJU International. 2007;100(2):351-6.

9. Klarskov N, Scholfield D, Soma K, Darekar A, Mills I, Lose G. Measurement of Urethral Closure Function in Women With Stress Urinary Incontinence. Journal of Urology. 2009;181(6):2628-33.

10. Mitchell PJ, Klarskov N, Hosker G, Lose G, Kiff ES. Anal acoustic reflectometry: a new technique for assessing anal sphincter function. Colorectal Disease. 2010;12(7):692-7.

11. Hornung BR, Mitchell PJ, Carlson GL, Klarskov N, Lose G, Kiff ES. Comparative study of anal acoustic reflectometry and anal manometry in the assessment of faecal incontinence. British Journal of Surgery. 2012;99(12):1718-24.

12. Klarskov N, Cerneus D, Sawyer W, Newgreen D, Van Till O, Lose G. The effect of single oral doses of duloxetine, reboxetine, and midodrine on the urethral pressure in healthy female subjects, using urethral pressure reflectometry. Neurourology and Urodynamics. 2018;37(1):244-9.

13. Klarskov N, Van Till O, Sawyer W, Cernus D, Sawyer W. Effect of a 5-HT2c receptor agonist on urethral closure mechanism in healthy women. Neurourology and Urodynamics. 2019;38(6):1700-6.

14. Zinner N, Scholfield D, Soma K, Darekar A, Grant L, Mills I. A Phase 2, 8-Week, Multi-Center, Randomized Double-Blind, Placebo-Controlled, Parallel Group Study Evaluating the Efficacy, Tolerability and Safety of [S,S] - Reboxetine (Pnu-165442g) for Stress Urinary Incontinence in Women. Journal of Urology. 2008;179(4S):569-70.

15. Kornholt J, Sonne DP, Riis T, Sonne J, Klarskov N. Effect of imipramine on urethral opening pressure: A randomized, double-blind, placebo-controlled crossover study in healthy women. Neurourology and Urodynamics. 2019;38(4):1076-80.

16. Christoffersen T, Kornholt J, Riis T, Sonne J, Sonne DP, Klarskov N. Effect of single doses of citalopram and reboxetine on urethral pressure: A randomized, double-blind, placebo- and active-controlled three-period crossover study in healthy women. Neurourology and Urodynamics. 2022;41(6):1482-8.

17. Collste L, Lindskog M. Phenylpropanolamine in treatment of female stress urinary incontinence Double-blind placebo controlled study in 24 patients. Urology. 1987;30(4):398-403.

18. Cardozo L, Lange R, Voss S, Beardsworth A, Manning M, Viktrup L, et al. Short- and long-term efficacy and safety of duloxetine in women with predominant stress urinary incontinence. Current Medical Research and Opinion. 2010;26(2):253-61.

19. Mitchell PJ, Klarskov N, Telford KJ, Hosker GL, Lose G, Kiff ES. Viscoelastic Assessment of Anal Canal Function Using Acoustic Reflectometry: A Clinically Useful Technique. Diseases of the Colon & Rectum. 2012;55(2):211-7.