Check for updates

Received: 15 October 2019 | Accepted: 30 December 2019

DOI: 10.1002/nau.24275

ORIGINAL CLINICAL ARTICLE

Neurourology Urodynamics ICS SUFU WILEY

Do overactive bladder symptoms and their treatment-associated changes exhibit a normal distribution? Implications for analysis and reporting

Marjan Amiri[1,2] | Sandra Murgas[3] | Andreas Stang[1,4] | Martin C. Michel[5]

[1]Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Essen, Germany

[2]Center for Clinical Trials Essen (ZKSE), University Hospital Essen, Essen, Germany

[3]Apogepha Arzneimittel GmbH, Dresden, Germany

[4]Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts

[5]Department of Pharmacology, Johannes Gutenberg University, Mainz, Germany

Correspondence
Martin C. Michel, Department of Pharmacology, Johannes Gutenberg University, Langenbeckstr. 1, Geb. 708, 1. OG, 55131 Mainz, Germany.
Email: marmiche@uni-mainz.de

Funding information
Apogepha; Innovative Medicines Initiative 2 Joint Undertaking, Grant/Award Number: 777364

Abstract

Aims: To explore the use of means vs medians (assuming or not the presence of normal distribution) in studies reporting overactive bladder syndrome symptoms and to test for normal distribution of basal values and treatment-associated changes thereof in two large noninterventional studies.

Methods: Systematic review of all original studies reporting on at least one overactive bladder syndrome symptom published in four leading urology journals in 2016 to 2017. Testing of the normal distribution of urgency, incontinence, frequency, and nocturia in two large noninterventional studies (n = 1335 and 745).

Results: Among 48 eligible articles, 86% reported means (assuming a normal distribution), 6% medians (not making this assumption), and 8% a combination thereof. Baseline values for all four symptoms and treatment-associated alterations thereof deviated from a normal distribution ($P < .0001$ in all cases). Means overestimated basal value and absolute changes thereof as compared with medians, for example, basal number of incontinence episodes in study 1 5.1 vs 4. Differences between means and medians for percentage changes of symptoms were small and did not consistently favor means over medians.

Conclusions: Dominant reporting of means implies the assumption of a normal distribution of overactive bladder syndrome symptoms but our data from two noninterventional studies do not support this assumption. We recommend that basal values and absolute symptom changes should be reported as medians and subjected to nonparametric analysis; means may be appropriate for the reporting of percentage changes of symptoms.

KEYWORDS
data analysis, normal distribution, overactive bladder syndrome, propiverine

1 | INTRODUCTION

The overactive bladder syndrome (OAB) has a high prevalence[1] and adversely impacts on the well-being of

Abbreviations: IQR, interquartile range; NIS, noninterventional study; OAB, overactive bladder syndrome; RCT, randomized controlled trial; SD, standard deviation.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2020 The Authors. *Neurourology and Urodynamics* Published by Wiley Periodicals, Inc.

the afflicted patients[2] and their family members.[3] Studies on OAB are mostly based on quantification of the four key symptoms of urgency, incontinence, frequency, and nocturia,[4] which typically are captured from voiding diaries. When looking at published OAB studies, we noticed that some investigators report baseline symptoms and treatment-associated changes thereof as means, whereas others report them as medians. The use of means appears intuitive to many, but it assumes that the data come from a population exhibiting a normal (Gaussian) distribution. A normal distribution is characterized by a unimodal, symmetrical distribution with ~68% of all values falling within ± 1 standard deviation (SD). In contrast, medians provide a useful description of the central tendency of a unimodal distribution that is not normal. For distributions that have more than one peak (multimodal distributions), both measures of central tendency are inappropriate. Whether a normal distribution exists also has implications on the choice of reported error bars and statistical tests. The SD and parametric tests such as *t* tests are only appropriate if a normal distribution can be assumed.

Against this background, we have systematically extracted information from four major urology journals reporting on OAB studies to explore whether they reported means or medians of OAB symptoms and whether they provided a justification for this choice. Moreover, we have used data from two large noninterventional studies (NIS) in OAB treatment to determine whether it is justified to assume a normal distribution. We have compared means and medians to determine whether using one yields a systematic over- or underestimation as compared with the other. Such calculations were made for the four OAB symptoms of urgency, incontinence, frequency, and nocturia and for treatment-associated alterations thereof; potential differences in reporting between studies with and without industry involvement was a secondary exploratory aim of our analysis.

2 | PATIENTS AND METHODS

2.1 | Present reporting practice in peer-reviewed publications

We conducted a systematic review of the original studies published in 2016 and 2017 in four major urology journals (BJU International, European Urology, Journal of Urology, Neurourology, and Urodynamics) that reported on at least one of the OAB symptoms of urgency, incontinence, frequency, and nocturia. We used the PubMed database (www.ncbi.nlm.nih.gov/pubmed) to identify relevant studies. In our PubMed search performed in May and June 2018, we entered the "name of journal" and "overactive bladder" in the search field and applied the date filter 2016.01.01-2017.12.31 for either print or online publication. To minimize selection bias, we included all original studies written in English and reporting on at least one of the four OAB symptoms within this period, including those that did not study OAB. We included randomized and observational studies investigating a medical or a surgical intervention including implanting devices as well as studies comparing groups of patients without studying an intervention. Our systemic review excluded preclinical studies, reviews, editorials, and letters. However, we did not use any filter regarding the sample size of the reported study. MA extracted the studies' primary and secondary endpoints and their respective measurement methods from the full text of each article. Moreover, she checked whether the authors provided a justification for their choice of reporting means or medians. In addition, she extracted the authors' affiliations and study funding sources to explore whether a study had involved industry. Individual studies considered in our analysis and a PRISMA flow chart are listed in the Online Supporting Information.

2.2 | Clinical data from two NIS

We used data from two NIS of a similar design and performed in 2012 and 2014 for a post hoc analysis. They included 1335 and 745 patients, respectively, and hereafter named study 1 and 2. Both studies had been performed with approval from the ethical committee of the state board of physicians in Saxony, Germany (Sächsische Landesärztekammer EK-BR-14/12-1 and EK-BR-18/14-1). While the analyses occurred after the studies had been completed, the statistical analysis plan had been prespecified before any analysis related to the normality of data distribution. Both NIS asked participating physicians to document baseline data and treatment outcomes for patients receiving propiverine ER based on the physician's judgment to treat their OAB symptoms. The planned duration of observation was 12 weeks with planned visits at baseline and after ~4 and 12 weeks. According to the applicable prescribing information, the starting dose could be 30 mg or 45 mg once daily and could be adapted during the duration of the studies. Data were collected on standardized case record forms. Based on study protocol, OAB symptom intensity assessment was based on voiding diaries but in line with the noninterventional character of the study, the length of observation period per assessment period in the diary was left to the discretion of the physician.

Our analysis of the baseline data was based on all patients having an entry for a given parameter. Patients not exhibiting a given symptom at baseline (value of 0) and those with medically implausible values (urgency > 50, frequency > 40,

nocturia > 20, and incontinence > 30) were excluded from the analysis for that symptom; this affected four patients each for urgency and frequency, one for nocturia, and one for incontinence in study 1 and none in study 2. Our analysis of the treatment data was based on a subgroup of the baseline cohort: to minimize heterogeneity based on dosing decision and duration of follow-up, this included only patients with a starting dose of 30 mg, having a recorded value at the 12-week time point and no change of administered dose during the observation period. We chose this group because it represents the majority of patients. We evaluated treatment-associated reductions in symptom frequency effects as delta (baseline – 12-week value) and as percentage reduction; for mathematical reasons, we calculated the latter only for patients with a baseline value other than 0 for the respective parameter. Missing data for one parameter did not exclude the use of other parameters from the same patient.

Based upon a reviewer's suggestion, we have performed a post hoc analysis to determine whether professional statisticians had been involved in the analysis of the data of the published papers. For this purpose, we checked the published manuscripts whether any author listed an affiliation to a statistics, biostatistics, or epidemiology department. Furthermore, we contacted each corresponding author to ask for the involvement of a professional statistician. If either was positive, we assumed the involvement of a statistician. Moreover, we asked corresponding authors whether to their knowledge a professional statistician had been involved as part of the manuscript evaluation by the journal. For either assessment, we compared numbers of articles reporting means, medians, or a combination thereof in an exploratory manner.

We tested for normal distribution using the D'Agostino and Pearson K2 omnibus test. To assess the impact of a lack of normal distribution, we compared means and medians. We performed all data analysis using the Prism software (version 8.2.1; GraphPad, La Jolla, CA). As we report only on a subset of both clinical studies, full data including those on tolerability will be presented in a subsequent report.

3 | RESULTS

3.1 | Present reporting practice in peer-reviewed publications

We retrieved and reviewed 183 articles in total, 16, 27, 39, and 101 articles from BJU International, European Urology, Journal of Urology, Neurourology, and Urodynamics, respectively (see Online Supporting Information). Forty-eight papers were eligible for inclusion in our analysis. Most of the articles (86%) reported means, only a few medians (6%) and some both means and medians (8%). All studies with industry involvement reported means; among those without industry involvement, corresponding numbers were 75% means, 11% medians, and 14% a combination thereof. The articles reporting means typically showed SD, standard error, or confidence interval error bars, which were internally consistent in their assumptions of a normal distribution. In contrast, those reporting medians showed error bars as interquartile ranges (IQR), which were also internally consistent in not assuming a normal distribution.

Only four studies (8%) provided information on testing for normal distribution within their data set: two studies reported having used the Shapiro-Wilk's test[5,6] and one the Kolmogorov-Smirnov test and a distribution histogram[7]; they did not disclose the results of normality testing but presented data as medians. One study showed a histogram for end-of-treatment values that clearly showed a lack of normal distribution but nonetheless reported means.[8] Of note, some of these studies included only 58[6] and 132 subjects,[7] indicating that they were probably too small to allow robust conclusions on the normal distribution. Four other studies (8%) made statements on normality but did not mention on which analysis this was based. One of them claimed to have applied parametric and nonparametric tests to parameters with and without normal distribution, respectively but did not disclose which applied to which parameter and concomitantly reported means and medians for all OAB symptoms.[9] Three others claimed normal distribution without showing supporting data and reported mean values.[10–12] The studies providing some justification for the choice of means vs medians reported both baseline symptoms and treatment responses but did not differentiate their assumptions related to a normal distribution for the two assessments.

In a post hoc analysis, we explored the impact of the involvement of a professional statistician in the data analysis on reporting. Twenty-one of 26 papers with available information had involved a statistician; they reported means in 16, medians in three and a combination thereof in two cases. Five of 26 papers reported not having involved a statistician; they all reported means. Of the 22 papers without information about the involvement of a statistician, 20 reported means and two a combination of means and medians. Authors of eight papers reported that the referee comments included specific feedback on statistical analysis; the published papers reported means in six and a combination of means and medians in two cases. Responding authors of four papers stated that no specific statistical review was provided; their published papers all reported means.

TABLE 1 Baseline severity of OAB symptoms (episodes per 24 hours)

	Study 1			Study 2		
Symptom	**n**	**Mean ± SD**	**Median (IQR)**	**n**	**Mean ± SD**	**Median (IQR)**
Urgency	1151	10.5 ± 5.9	10 (6; 14)	621	10.0 ± 5.5	10 (6; 13)
Incontinence	785	5.1 ± 3.9	4 (2; 7)	418	5.5 ± 3.9	5 (2; 7)
Frequency	1308	13.6 ± 4.4	13 (11; 16)	730	13.2 ± 4.2	13 (10; 15)
Nocturia	1269	3.4 ± 1.6	3 (2; 4)	706	3.5 ± 1.7	3 (2; 4)

Note: Patients not exhibiting urgency, incontinence, or nocturia were excluded for that parameter. All four parameters differed from a normal distribution in the D'Agostino and Pearson K2 omnibus test at $P < .0001$.
Abbreviations: IQR, interquartile range; OAB, overactive bladder syndrome.

3.2 | Clinical data from two NIS

The distribution of symptom intensity (episode frequency) differed significantly from normality for all four OAB symptoms in both studies ($P < .0001$ for all parameters; Table 1). As an example, a graphical representation of the distribution of intensity based on frequency data is shown as Figure 1 indicating that the data exhibited a unimodal, but not symmetrical distribution. Accordingly, means were systematically higher than the corresponding medians in both studies. The means as surrogate values for medians overestimated the medians in the two data sets by 0.5 and 0.0 for urgency, 1.1 and 0.5 for incontinence, 0.6 and 0.2 for frequency, and 0.4 and 0.5 for nocturia.

The distribution of treatment responses expressed as absolute changes of episode frequency differed significantly from normality in both NIS ($P < .0001$ for all parameters; Table 2). Like baseline values (see above), means as surrogate measures for the more appropriate medians of absolute differences systematically overestimated treatment effects on urgency (1.1 and 0.5), incontinence (0.7 and 0.9), and frequency (0.8 and 0.8) in both studies; however, there was no overestimation of changes in nocturia episodes (0 and −0.3).

The distribution of treatment responses expressed as the relative difference in episode frequency (percentage of week 12 related to baseline measurement) also differed significantly from normality in both NIS ($P < .0001$ for all parameters; Table 3). However, differences were small and, if anything, means underestimated symptom changes relative to medians for all four symptoms.

4 | DISCUSSION

4.1 | Critique of methods

Our systematic review was based on four leading journals in the OAB field. These journals were chosen assuming that leading journals may have published papers of the highest applicable standards. To minimize selection bias, the chosen years of publication were systematically screened for each journal.

Our analysis of clinical data was based on two NIS, not on randomized controlled trials (RCT). This was done because NIS tend to be larger than RCT, and a robust

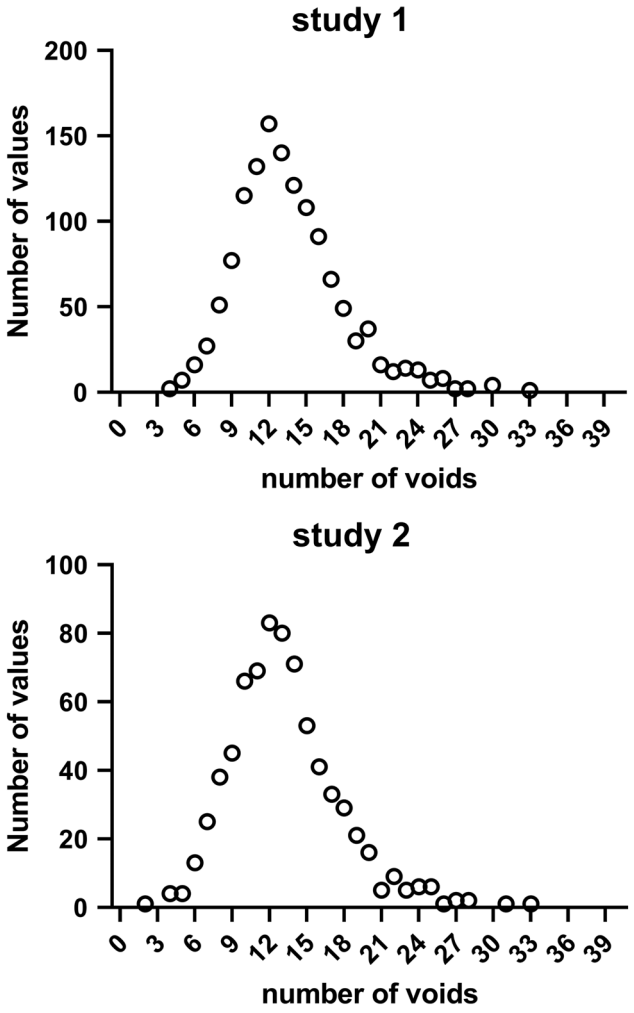

FIGURE 1 Distribution of basal micturition frequency (number of voids/24 hours) in studies 1 and 2. The highest observed values in both studies were 33

TABLE 2 Absolute reductions of OAB symptoms (delta of episodes per 24 hour) after 12 weeks of treatment in the subgroup continuously receiving 30 mg/day

	Study 1			Study 2		
Symptom	**n**	**Mean ± SD**	**Median (IQR)**	**N**	**Mean ± SD**	**Median (IQR)**
Urgency	627	7.1 ± 5.2	6 (3; 10)	335	6.5 ± 4.6	6 (3; 9)
Incontinence	414	3.7 ± 3.0	3 (2; 5)	218	3.9 ± 3.3	3 (1; 5)
Frequency	740	5.8 ± 3.7	5 (3; 7)	415	4.8 ± 3.5	4 (3; 6)
Nocturia	727	2.0 ± 1.9	2 (1; 3)	401	1.7 ± 1.5	2 (1; 2)

Note: All four parameters differed from a normal distribution in the D'Agostino and Pearson K2 omnibus test at $P < .0001$.
Abbreviations: IQR, interquartile range; OAB, overactive bladder syndrome.

assessment of normality requires a large sample to be analyzed.

The data sets being analyzed had been collected as part of the obligation of pharmaceutical companies in Germany to provide ongoing collection on the safety and tolerability of a medicine (German Arzneimittelgesetz). While our data represent post hoc analyses, they were based on a statistical analysis plan that had been finalized before data were inspected related to the goals of our investigation to avoid bias during the analysis process. Several statistical tests are available to test for deviation from a normal distribution. We had selected the D'Agostino and Pearson K2 omnibus test as a primary outcome measure. However, three other tests for normality (Anderson-Darling, Shapiro-Wilk, and Kolmogorov-Smirnov) consistently confirmed the deviation from normality for each parameter in each study (data not shown).

Previous NIS on the use of muscarinic receptor antagonists in the treatment of OAB have typically reported mean values, for instance, based on treatment with darifenacin,[13] solifenacin,[14] and tolterodine.[15,16] Mean baseline intensity of symptoms and treatment-associated improvements thereof in the previous four and in the present two studies were comparable, indicating that we have used data sets that are representative for the overall population of patients with OAB seeking medical treatment in a real-life setting. This is in line with the general observation that all muscarinic receptor antagonists have comparable efficacy.[17,18] However, it was greater than observed in most RCT,[17,18] at least in part because RCT typically includes a single-blind placebo run-in period before establishing baseline symptoms. In conclusion, our clinical analyses are based on data sets comparable with those of many other NIS and, therefore, our clinical findings may be generalizable.

4.2 | Present reporting practice in peer-reviewed publications

Our results show that reporting of distributions of symptoms or changes of symptoms related to OAB are statistically inconsequently and therefore potentially misleadingly described. Most authors used means, implicitly assuming a normal distribution of OAB symptoms. However, most of the authors reporting means did not provide any justification for their assumption of a normal distribution. With one exception,[8] the few articles providing a justification for assuming a normal distribution did not reference or show the data supporting it. Although it can be assumed that pharmaceutical companies sponsoring a clinical trial have professional statisticians on staff, industry-sponsored studies consistently used means and failed to

TABLE 3 Relative reduction of OAB symptoms (percentage reduction of episodes per 24 hour) after 12 weeks of treatment in the subgroup continuously receiving 30 mg/day

	Study 1			Study 2		
Symptom	**n**	**Mean ± SD**	**Median (IQR)**	**n**	**Mean ± SD**	**Median (IQR)**
Urgency	627	71 ± 29	75 (56; 92)	335	65 ± 30	67 (50; 90)
Incontinence	414	82 ± 30	100 (67; 100)	218	70 ± 39	78 (50; 100)
Frequency	740	41 ± 17	42 (30; 50)	415	36 ± 20	36 (25; 46)
Nocturia	727	59 ± 29	60 (50; 75)	401	48 ± 39	50 (33; 67)

Note: All four parameters differed from a normal distribution in the D'Agostino and Pearson K2 omnibus test at $P < .0001$.
Abbreviations: IQR, interquartile range; OAB, overactive bladder syndrome.

provide a justification for this. The involvement of a professional statistician was associated with the choice of measures of central tendency (mean or median): while medians were reported only when a statistician was involved, studies with such involvement in most cases also reported means. When a statistical evaluation had apparently been part of the peer review process of a manuscript, the proportion of publications containing medians was somewhat higher although the majority of reports still were limited to means. Based on these post hoc analyses, it appears that the involvement of a statistician in the data analysis and/or peer review of the manuscript made it more likely that medians were reported but even in those cases means were more common.

The heterogeneity in reporting of OAB parameters and the lack of providing data underlying it (except one of 48 studies) shows that an analysis to determine the validity of the assumption of a normal distribution is necessary. For variables that can only provide non-negative values, a simple heuristic makes clear that reporting means and SD is inappropriate: if the mean minus 1 to 2 SD's predicts negative values for the 95% range of the data, the assumption of normality cannot be correct as the descriptive statistic predicts values that are impossible (eg, present study 1: urgency: mean 10.7 and SD 6.6).

4.3 | Clinical data from two NIS

When a given symptom at baseline does not exhibit a normal distribution, the absolute (after minus before treatment) or relative (%) difference of symptoms can still exhibit a normal distribution. Our systematic literature review revealed only one study that showed a non-normal distribution of the treatment responses assessed as difference[8] but did not disclose similar data at baseline. Our analysis consistently found evidence of a deviation from normality across two large data sets, four OAB symptoms and for baseline data and treatment responses. This is in line with the limited data from other investigators focusing on treatment responses only.[8] Therefore, we conclude that it cannot necessarily be assumed that OAB diary data and their improvement upon treatment exhibit a normal distribution in the general population. It follows, that the distribution of OAB symptoms has to be checked carefully before authors can decide which measure of central tendency (means vs medians) and spread (SD vs IQR) are reasonable.

A true normal distribution is rare for any parameter in biomedical research. Therefore, the more relevant question is whether the extent of deviation from normality is large enough to make use of means and parametric null hypothesis tests assuming such normality misleading. To explore this, we have compared the means and medians of baseline values and treatment responses for each parameter in each study. According to our data, the difference between mean and median baseline values was 0.5 and 0.0 for urgency, 1.1 and 0.5 for incontinence, 0.6 and 0.2 frequency, and 0.5 and 0.5 for nocturia. For absolute treatment effects, that is, the difference between the number of symptoms, it was 1.1 and 0.5 for urgency, 0.7 and 0.9 for incontinence, 0.8 and 0.8 for frequency, and 0 and −0.3 for nocturia. These differences are clinically relevant because they are comparable with the difference in treatment effects between muscarinic antagonists and placebo in the reduction of incontinence and frequency episodes, which according to meta-analyses are < 1 and < 1.5 per day, respectively.[18] They are also comparable with reported differences between the β_3-adrenoceptor agonist mirabegron and placebo for a number of incontinence or micturition episodes (0.44 and 0.62, respectively)[19] and for comparisons between active treatments.[17] Similarly, the minimum noticeable change in incontinence episodes as assessed by patients using a quality of life tool was reported to be 3 per week,[20] that is, about 0.43 episodes/24 hours. The differences between mean and median in baseline or treatment-induced change of incontinence episodes in our two studies exceeded this threshold for being detectable by patients. Therefore, we conclude that the unsupported assumption of a normal distribution of OAB symptoms is not only theoretically flawed but also leads to an overestimation of symptom intensity and treatment improvements that are comparable with or greater than the placebo-corrected effect of muscarinic antagonists and greater than what has been reported to be noticeable by patients.

A different situation may exist for relative treatment effects (the difference between the number of symptoms after and before the therapy, divided by the number of symptoms before the therapy) expressed in percent. Although percentage changes also exhibited a deviation from a normal distribution, the resulting differences between means and medians typically were smaller and, if anything, means of percentage changes underestimated treatment effects. These differences most likely are not clinically relevant when compared with differences between active treatment and placebo[18,19] or to minimal differences noticeable by patients.[20] Therefore, it appears justifiable to report percentage changes of OAB symptoms as means. Some studies report baseline-adjusted absolute improvements of treatment,[21,22] which is conceptually the same as percentage improvements of treatment effects. Therefore, means of baseline-adjusted treatment effects may also be an acceptable way of

reporting efficacy data despite the formal deviation from the normal distribution. In a more general vein, not making assumptions on normal distribution in the absence of robust supporting data is the safer option for analysis of data.

One may expect that regulatory authorities such as the European Medicines Agency or the US Food and Drug Administration with their vast knowledge of treatment studies would have issued guidance on the use of means vs medians. However, the available guidance documents from both agencies related to OAB do not address the use of means vs medians or parametric vs nonparametric statistical analysis. However, the European Medicines Agency has issued general guidance that data should be checked for normality of distribution of reported variables and that analysis and presentation of the data should be based on this.[23]

5 | CONCLUSIONS

We conclude that most investigators report means of OAB symptoms and treatment-induced changes thereof, implying the assumption that these parameters exhibit a normal distribution. Our data shows that this assumption and the reporting of means may be scientifically unjustified and may potentially result in misinterpretations of study results on OAB symptoms. The difference between means and medians for OAB symptoms and symptom differences is likely to be of clinical relevance. Relative improvements of OAB symptoms (percentages changes) may be an exception from this, that is, even if there is no true normal distribution the consequences for parameter estimates are minimal. While our data are based on the use of propiverine, we propose that they may also be applicable to other treatments including muscarinic receptor antagonists as a class, β_3-adrenoceptor agonists, onabotulinum toxin A or behavioral or other nonpharmaceutical interventions.

ACKNOWLEDGMENT

This study is supported by Apogepha (the employer of authors Sandra Murgas) and Innovative Medicines Initiative 2 Joint Undertaking (grant no. 777364); this Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA.

ORCID

Andreas Stang http://orcid.org/0000-0001-6363-9061
Martin C. Michel http://orcid.org/0000-0003-4161-8467

REFERENCES

1. Coyne KS, Sexton CC, Thompson CL, et al. The prevalence of lower urinary tract symptoms (LUTS) in the USA, the UK and Sweden: results from the epidemiology of LUTS (EpiLUTS) study. *BJU Int*. 2009;104:352-360.
2. Coyne KS, Kvasz M, Ireland AM, Milsom I, Kopp ZS, Chapple CR. Urinary incontinence and its relationship to mental health and health-related quality of life in men and women in Sweden, the United Kingdom, and the United States. *Eur Urol*. 2012;61:88-95.
3. Coyne KS, Matza LS, Brewster-Jordan J, Thompson C, Bavendam T. The psychometric validation of the OAB family impact measure (OAB-FIM). *Neurourol Urodyn*. 2010;29:359-369.
4. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the standardisation sub-committee of the International Continence Society. *Neurourol Urodyn*. 2002;21:167-178.
5. Kubota Y, Hamakawa T, Osaga S, et al. A kit ligand, stem cell factor as a possible mediator inducing overactive bladder. *Neurourol Urodyn*. 2018;37:1258-1265.
6. Blais A-S, Nadeau G, Moore K, Genois L, Bolduc S. Prospective pilot study of mirabegron in pediatric patients with overactive bladder. *Eur Urol*. 2016;70:9-13.
7. Azuri J, Kafri R, Ziv-Baran T, Stav K. Outcomes of different protocols of pelvic floor physical therapy and anti-cholinergics in women with wet over-active bladder: a 4-year follow-up. *Neurourol Urodyn*. 2017;36:755-758.
8. Martina R, Kay R, Abrams P, van Maanen R, Ridder A. A clinical perspective on the analysis and presentation of the number of incontinence episodes following treatment for OAB. *Neurourol Urodyn*. 2016;35:728-732.
9. Abulseoud A, Moussa A, Abdelfattah G, Ibrahim I, Saba E, Hassouna M. Transcutaneous posterior tibial nerve electrostimulation with low dose trospium chloride: could it be used as a second line treatment of overactive bladder in females. *Neurourol Urodyn*. 2018;37:842-848.
10. Drake MJ, Chapple C, Esen AA, et al. Efficacy and safety of mirabegon add-on therapy to solifenacin in incontinent overactive bladder patients with an iInadequate response to initial 4-week solifenacin monotherapy: a randomised double-blind multicentre phase 3B study (BESIDE). *Eur Urol*. 2016;70:136-145.
11. Borch L, Hagstroem S, Kamperis K, Siggaard CV, Rittig S. Transcutaneous electrical nerve stimulation combined with oxybutynin is superior to monotherapy in children with urge incontinence: a randomized, placebo controlled study. *J Urol*. 2017;198:430-435.
12. Koschorke M, Leitner L, Sadri H, Knüpfer SC, Mehnert U, Kessler TM. Intradetrusor onabotulinumtoxinA injections for refractory neurogenic detrusor overactivity incontinence: do we need urodynamic investigation for outcome assessment? *BJU Int*. 2017;120:848-854.
13. Schneider T, Marschall-Kehrel D, Hanisch JU, Michel MC. Do gender, age or life style factors affect responses to antimuscarinic treatment in overactive bladder patients? *Int J Clin Pract*. 2010;64:1287-1293.
14. Witte LPW, Peschers U, Vogel M, de la Rosette JJMCH, Michel MC. Does the number of previous vaginal deliveries affect overactive bladder symptoms or their response to treatment? *LUTS*. 2009;1:82-87.

15. Michel MC, Schneider T, Krege S, Goepel M. Do gender or age affect the efficacy and safety of tolterodine? *J Urol*. 2002;168:1027-1031.
16. Michel MC, de la Rosette JJ, Piro M, Schneider T. Comparison of symptom severity and treatment response in patients with incontinent and continent overactive bladder. *Eur Urol*. 2005;48:110-115.
17. Novara G, Galfano A, Secco S, et al. A systematic review and meta-analysis of randomized controlled trials with antimuscarinic drugs for overactive bladder. *Eur Urol*. 2008;54:740-764.
18. Reynolds WS, McPheeters M, Blume J, et al. Comparative effectiveness of anticholinergic therapy for overactive bladder in women. A systematic review and meta-analysis. *Obstet Gynecol*. 2015;125:1423-1432.
19. Cui Y, Zong H, Yang C, Yan H, Zhang Y. The efficacy and safety of mirabegron in treating OAB: a systematic review and meta-analysis of phase III studies. *Int Urol Nephrol*. 2014;46:275-284.
20. Homma Y, Koyama N. Minimally clinically important change in urinary incontinence detected by a quality of life assessment tool in overactive bladder syndrome with urge incontinence. *Neurourol Urodyn*. 2006;25:228-235.
21. Wagg A, Dale M, Tretter R, Stow B, Compion G. Randomised, multicenre, placebo-controlled, double-blind crossover study investigating the effect of solifenacin and oxybutynin in elderly people with mild cognitive impairment: the SENIOR study. *Eur Urol*. 2013;64:74-81.
22. Chapple CR, Dvorak V, Radziszewski P, et al. A phase II dose-ranging study of mirabegron in patients with overactive bladder. *Int Urogynecol J*. 2013;24:1447-1458.
23. European Medicines Agency. ICH Topic E9. Statistical principles for clinical trials; 1998. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf. Accessed December 5, 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Amiri M, Murgas S, Stang A, Michel MC. Do overactive bladder symptoms and their treatment-associated changes exhibit a normal distribution? Implications for analysis and reporting. *Neurourology and Urodynamics*. 2020;39:754–761. https://doi.org/10.1002/nau.24275