



Von Frey testing revisited: Provision of an online algorithm for improved accuracy of 50% thresholds

Christensen, Sarah Louise; Hansen, Rie Bager; Storm, Mikkel Ammentorp; Olesen, Jes; Hansen, Thomas Folkmann; Ossipov, Michael; Izarzugaza, Jose M. G.; Porreca, Frank; Kristensen, David M.

Published in:
European Journal of Pain

DOI:
[10.1002/ejp.1528](https://doi.org/10.1002/ejp.1528)

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Christensen, S. L., Hansen, R. B., Storm, M. A., Olesen, J., Hansen, T. F., Ossipov, M., ... Kristensen, D. M. (2020). Von Frey testing revisited: Provision of an online algorithm for improved accuracy of 50% thresholds. *European Journal of Pain*, 24(4), 783-790. <https://doi.org/10.1002/ejp.1528>

Von Frey testing revisited: Provision of an online algorithm for improved accuracy of 50% thresholds

Sarah Louise Christensen¹ | Rie Bager Hansen² | Mikkel Ammentorp Storm¹ |
 Jes Olesen¹ | Thomas Folkmann Hansen¹ | Michael Ossipov^{3,4} | Jose M. G. Izarzugaza⁵ |
 Frank Porreca⁴ | David M. Kristensen^{1,6}

¹Danish Headache Center, Glostrup Research Institute, Rigshospitalet Glostrup, Denmark

²Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark

³Syneos Health, Raleigh, NC, USA

⁴Department of Pharmacology, College of Medicine, University of Arizona Health Sciences Center, Tucson, AZ, United States

⁵Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark

⁶University of Rennes, Inserm, EHESP, Irset (Institut de Recherche en Santé, Environnement et Travail) – UMR_S 1085, Rennes, France

Correspondence

Sarah Louise Christensen, Danish Headache Center and Glostrup Research Institute, Rigshospitalet Glostrup, Glostrup, Denmark.

Email: schr0227@regionh.dk; sarah.tangsgaard@gmail.com

Funding Information

The work was supported by Candys Foundation. The funder had no involvement in the study.

Abstract

Background: In the pain field, it is essential to quantify nociceptive responses. The response to the application of von Frey filaments to the skin measures tactile sensitivity and is a surrogate marker of allodynia in states of peripheral and/or central sensitization. The method is widely used across species within the pain field. However, uncertainties appear to exist regarding the appropriate method for analysing obtained data. Therefore, there is a need for refinement of the calculations for transformation of raw data to quantifiable data.

Methods: Here, we briefly review the fundamentals behind von Frey testing using the standard up-down method and the associated statistics and show how different parameters of the statistical equation influence the calculated 50% threshold results. We discuss how to obtain the most accurate estimations in a given experimental setting.

Results: To enhance accuracy and reproducibility across laboratories, we present an easy to use algorithm that calculates 50% thresholds based on the exact filaments and their interval using math beyond the traditional methods. This tool is available to the everyday user of von Frey filaments and allows the insertion of all imaginable ranges of filaments and is thus applicable to data derived in any species.

Conclusion: We advocate for the use of this algorithm to minimize inaccuracies and to improve internal and external reproducibility.

Significance: The von Frey testing procedure is standard for assessing peripheral and central sensitization but is associated with inaccuracies and lack of transparency in the associated math. Here, we describe these problems and present a novel statistical algorithm that calculates the exact thresholds using math beyond the traditional methods. The online platform is transparent, free of charge and easy to use also for the everyday user of von Frey filaments. Application of this resource will ultimately reduce errors due to methodological misinterpretations and increase reproducibility across laboratories.

1 | INTRODUCTION

Evaluation of cutaneous sensitivity to mechanical stimulation with von Frey filaments as a surrogate marker of peripheral and/or central sensitization continues to be standard in both preclinical and human pain research (Bradman, Ferrini, Salio, & Merighi, 2015). Since its conception in the late 19th century, several methodologies for the actual testing paradigm and subsequent conversion into pain thresholds have been developed. The most common of these is the “up-down method” for approximation of the 50% withdrawal threshold aiming to determine the weight of stimulus that elicits a response 50% of the times it is applied (Mills et al., 2012). Dixon provided the statistical framework for these types of calculations (Dixon, 1965, 1980) and Chaplan and colleagues applied his method in a testing paradigm for evaluation of tactile sensitivity thresholds in rats (Chaplan, Bach, Pogrel, Chung, & Yaksh, 1994).

The approach is not without caveats. One often-ignored issue is that the method was originally developed for rats and that the shift in weight of filaments for other species can cause irregularities in essential factors of the equation for calculation of 50% thresholds. The equation is often used without fully comprehending the mathematics behind the work of Dixon (1965, 1980) resulting in different outcomes and unnecessary variation. These issues have previously been addressed and solutions to some of the irregularities proposed (Bradman et al., 2015; Gonzalez-Cano et al., 2018; Milligan et al., 2000; Mills et al., 2012). Unfortunately, there is still problems and variation in the use of the Dixon–Chaplan method.

In this article, we show how the different parameters of the Dixon statistics in Chaplan's equation may influence the estimation of 50% withdrawal thresholds and provide a method to obtain the most accurate data in a given experimental setting. We provide an online calculator that offers an easy to use, versatile algorithm to determine 50% withdrawal thresholds applying computational power to improve the calculations. Use of this resource will reduce the inaccuracies associated with the traditional use of Chaplan's equation.

2 | METHODS

2.1 | Von Frey filaments and the up-down method

Von Frey (Semmes–Weinstein) filaments are calibrated nylon mono filaments that apply different forces to the stimulated skin. Individual filaments are labelled by a dimensionless handle number (e.g. 4.31 or 3.61) and a target force measured in grams (e.g. 2.0 or 0.4 g). At this point, we should note that the term “grams” is commonly used to describe force, in place of the more strictly proper “newtons.” However, since the conversion of grams to newtons for all filaments would

be done by an identical multiplier, this technically inaccurate term is not a factor. The handle number and the strength (in grams) of the filament are related, where the handle number is the \log_{10} of the force expressed as 1/10th mg. This is done simply to avoid handles with negative numbers. This happens because the monofilaments were originally designed to vary by \log_{10} increments of 1/10 mg (Bradman et al., 2015):

$$A) \text{ Weight, } g = 10^{\text{Handle number}} / 10,000$$

However, studies where the force applied by the commonly used Semmes–Weinstein monofilaments were measured and compared to the target force as calculated by the given equation revealed that there are some deviations (Bradman et al., 2015), especially for the lightest filaments; see Table 1. Importantly, even the Operation Manual supplied by Stoelting for its Semmes–Weinstein monofilament kit explicitly state that the “calculated and actual forces do not always agree.” Hence, we recommend the use of target force for calculation and reference given its increased accuracy. This will be further exemplified later. Ideally, each set of Semmes–Weinstein monofilaments should be calibrated individually and rechecked over time to detect possible changes in strength after repeated use (Bradman et al., 2015; Werner, Rotbøll-Nielsen, & Ellehuus-Hilmersson, 2011).

The range of filaments to be used in a specific experiment is determined according to the given situation. For illustrative purposes, we will use the following range of filaments labelled by force and handle number in parenthesis: 0.008 g (1.65), 0.02 g (2.36), 0.04 g (2.44), 0.07 g (2.83), 0.16 g (3.22), 0.4 g (3.61), 0.6 g (3.84), 1.0 g (4.08) and 2.0 g (4.31). These are based on testing of the plantar surface of the hind paw of mice. Here, almost all naïve mice respond to the application of the 2.0 g filament (4.31) this is the upper limit of the testing range. The lowest filament is 0.008 g (1.65) to which only extremely sensitive mice respond. In practice, it may be necessary to shift the scale towards heavier filaments if naïve mice do not respond to 2 g. The converse would be necessary for testing highly allodynic areas or delicate body regions (Burgos-Vega et al., 2018). The 1.4 g (4.17) filament is not included in the range to avoid the small interval obtained between 4.08 and 4.17. This interval deviates considerably from the mean interval of the entire range (Bradman et al., 2015); see the section on determination of δ .

2.2 | Determination of 50% withdrawal thresholds by the up-down (Chaplan) method

The up-down method is a statistical tool for determination of 50% thresholds based on a sequential array of experiments (Dixon, 1965, 1980). The basis of estimating the 50% threshold is the binominal pattern of positive and negative responses to different stimulations, in this case achieved with von Frey

TABLE 1 Deviations between handle number and target force

Handle number	Target force, g	Calculated force, g	Deviation, %
1.65	0.008	0.005	-44.2
2.36	0.02	0.02	14.5
2.44	0.04	0.03	-31.1
2.83	0.07	0.07	-3.4
3.22	0.16	0.18	3.7
3.61	0.4	0.41	1.9
3.84	0.6	0.69	15.3
4.08	1.0	1.2	20.2
4.17	1.4	1.47	5.7
4.31	2.0	2.04	2.1
4.56	4.0	3.6	-9.2
4.74	6.0	5.5	-8.4
4.93	8.0	8.5	6.4
5.07	10	11.7	17.5
5.18	15	15.1	0.9

Note: Mathematically, handle number and target force are related, equation A. However, the historical characteristics of handle number give rise to deviations from its corresponding force. First two columns; handle number and corresponding target force as provided by the manufacture of filaments. Third column; force as calculated from handle number using formula A. Last column; per cent-wise deviation of each filament from its own target force if handle number is applied to recognize the filaments.

filaments. Accordingly, using the range of filaments described above we would first apply the middle weight 0.16 g (3.22). With a negative response (O) stimulus is increased one step and with a positive response (X) the stimulus is decreased. Following the breaking point (XO/OX) another four stimuli are applied resulting in a certain pattern of negative and positive responses, for example, OOOXOXOX, XXXOXOXO or OOXOXOX (Chaplan et al., 1994; Dixon, 1965, 1980). Individual patterns have a corresponding constant, k, provided in a reference table by Dixon (Dixon, 1980). Chaplan and colleagues (Chaplan et al., 1994) adapted the statistical insights of Dixon to an equation for calculation of 50% withdrawal thresholds following stimulation with von Frey filaments:

$$B) 50\% \text{ threshold, } g = \left(10^{[X_r + k\delta]} \right) / 10,000$$

The k value is provided by the Dixon statistics table and is a constant related to the up-down pattern of the individual test

subject. The k value is based on a negative log likelihood function, consisting of a sum of log-transformed cumulative normal distributions with variance 1. This is obtained from an idealized situation with equal step size. δ is the mean difference in log units between stimuli. X_f is the value (in log units) of the last von Frey filament applied. If the lowest or highest filament is reached with no breaking point, the animal is assigned the value of the lowest or highest filament, respectively, without reference to Dixon statistics that do not account for such scenarios (Chaplan et al., 1994). In the following sections, we will describe the δ and X_f values and how these can give rise to inaccuracies in the estimation of the 50% withdrawal threshold.

2.3 | Biological ranking of XO response patterns without the use of Chaplan's formula

To compare different calculation methods for 50% withdrawal thresholds, different XO patterns were ranked in

TABLE 2 Frequency count of responses to the individual filaments of the XO patterns. Pattern (a) is less sensitive than (b) as it has 0/3 reactions to the 1 g filament, whereas (b) has 1/3 reactions to 1 g. (c) and (d) are equally sensitive as their response frequencies and proportions are identical

XO pattern/filament	0.008 g	0.02 g	0.04 g	0.07 g	0.16 g	0.4 g	0.6 g	1 g	2 g
a) OOOXOXOX 2 g	0/0	0/0	0/0	0/0	0/1	0/1	0/1	0/3	3/3
b) OOOXOXOX 0.6 g	0/0	0/0	0/0	0/0	0/1	0/1	0/2	1/3	2/2
c) XOXOXOX 0.4 g	0/0	0/0	0/0	0/2	2/3	1/1	0/0	0/0	0/0
d) XOXOXOX 0.07 g	0/0	0/0	0/0	0/2	2/3	1/1	0/0	0/0	0/0

relation to each other. For some patterns, this was done unambiguously by attention to the final filament applied, but for other patterns it could not be done that easily. Hence, we counted the frequency of responses to individual filaments to assess which pattern represents the more sensitive animal. In Table 2, examples of such frequency counts are provided.

From here, we conclude that a) is less sensitive than b) and that c) and d) are equally sensitive as they display same frequency of responses to individual filaments.

2.4 | Theoretical dataset

A theoretical dataset was created to illustrate the applicability of the mathematical improvements suggested in this manuscript. The theoretical experiment tested the efficacy of treatment X and Y in reversing a state of hyperalgesia. The question is whether one treatment is superior to the other? The dataset was created with a group size of 12 (two identical sets of six mouse XO responses) for each treatment. The data were created to include both XO patterns with deviations from the biological ranking and patterns without (raw data are provided as supplementary material). 50% withdrawal thresholds were calculated by six different methods in accordance with our results. Treatment X and Y are compared both parametrically and non-parametrically representing an ongoing discussion of what is correct (Bradman et al., 2015; Mills et al., 2012). The impact of the different methodological options is illustrated with group sizes of 6 and 12.

3 | RESULTS

3.1 | How to define X_f from Chaplan's equation

The last filament applied in the XO series is referred to as X_f in Chaplan's equation (B) and can be inserted as either the handle number or $\log(\text{target force})$. The discrepancy between the two estimates (Table 1) gives rise to differences in the estimation of 50% thresholds. In the original work by Chaplan et al., the handle number was used as X_f . [Chaplan et al., 1994, and T. Yaksh by personal communication]. However, we argue that data obtained using $\log(\text{target force})$ is more accurate. This is exemplified in the section *Consequences of different approaches to X_f and δ* . X_f can be accurately determined in-house by calibrating the filaments at frequent intervals (Bradman et al., 2015). When $\log(\text{target force})$ is inserted as X_f in formula B, the division by 10,000 is left out (see the relationship between handle number and target force, equation A).

3.2 | Determination of δ

The interval between stimuli is a fundamental discrepancy between the original method developed by Dixon and the use of the method to measure withdrawal thresholds with von Frey filaments. Dixon assumed equal spacing between stimuli but allowed up to 50% deviation of single steps from the mean (Dixon, 1980). However, Von Frey filaments do not have arithmetically even sequential steps but increase on an approximately logarithmic scale, and the intervals are not always equal (Mills et al., 2012). For example, the commonly used Semmes–Weinstein filaments are not all spaced equally, based on handle numbers (i.e. \log_{10} values). Accordingly, Chaplan defined δ as the mean difference (in log scale) between steps of the filament range. Chaplan's work was based on rat data and the range of filaments used for rats, which includes higher filament strengths than those used for mice, where greater deviations are noted (as explained above). Thus, in the range of filaments used for experiments with rats, the individual steps between filaments conform to the conditions defined by Dixon; see supplementary Table 1 for exact values. In contrast, calculation of δ for the range of filaments typically applied to mice (0.008 g, 0.02 g, 0.04 g, 0.07 g, 0.16 g, 0.4 g, 0.6 g, 1 g, 1.4 g, 2 g) reveals that some intervals within this range deviate from the requirements imposed by Dixon's method, since two of eight intervals deviate more than 50% from the mean of all the intervals when the calculation is based on handle numbers; see supplementary Table 1. The consequence of uneven steps is an under- or overestimation of the estimated thresholds (Bradman et al., 2015). This problem can be solved by determining δ based on $\log(\text{target force})$; a manoeuvre that changes the mean delta value from 0.3325 to 0.2997 for the specific mouse range of filaments included here. This definition of δ will be referred to as “constant δ ” hereafter. Another approach to determine δ is to base the value on each generated XO pattern. We have used two different methods to do this: the “flexible δ ” and “exact δ .” The flexible δ is calculated as the mean interval between the filaments that were actually applied to the individual test subject without the need for complicated mathematics (Bradman et al., 2015). The “exact δ ” is calculated by computational tools. The calculation does not use estimation of idealized k and δ values, but employs all the specific information obtained for each animal by the given XO pattern and array of filaments applied based on the original likelihood function described by Dixon (1965, 1980). Here, instead of using the δ value as standard deviation for the cumulative normal distribution, the step size for each individual step is used. Then the likelihood calculation is done iteratively for each observation in the sequential trial. For positive responses, the upper tail value of the cumulative normal distribution is used, for negative the lower tail value is used. These likelihoods are log transformed and summed to produce the final likelihood. The function is optimized for the mean of the normal distributions producing the most

likely weight stimulation for the 50% threshold. Importantly, the method avoids the bias of assuming regular intervals between filaments and consequently the over- and underestimations associated with irregular intervals. We have developed an online tool for estimations of 50% thresholds with all three definitions of δ , constant, flexible and exact (default). The algorithm recognizes filaments by both target force (default) and handle number and can be found at https://bioapps.shinyapps.io/von_frey_app/.

3.3 | Consequences of different approaches to X_f and δ

To illustrate the problems associated with the direct application of the Chaplan equation to mouse testing, we have calculated threshold values for selected XO patterns and depicted them graphically in descending order from left to right in agreement with the methodology presented

in Table 2. The sequence of patterns does not represent a quantitative scale, but a qualitative rank that are quantified by Chaplans formula (and modified formulas) on the y-axis. Figure 1 provides an overview of 50% thresholds with the six possible combinations for estimates of X_f and δ . Left-hand panels are handle number + constant/flexible/exact δ and right-hand panels are target force + constant/flexible/exact δ . For improved resolution, patterns of the upper range (top panels) and lower range (bottom panels) are presented separately. Looking at the calculated thresholds for patterns 1) and 2) in Figure 1a there are inaccuracies with the estimations as reaction pattern 1) responds only to the 2.0 g filament, whereas pattern 2) also responds to 1.0 g. Hence, pattern 2) represents a more sensitive test subject than pattern 1) but is erroneously assigned a higher 50% threshold. The same situation appears for the identical XO pattern when the breaking point is one increment lower, response patterns 5) and 6) in panels a and b. With the use of handle number and the constant δ the Chaplan

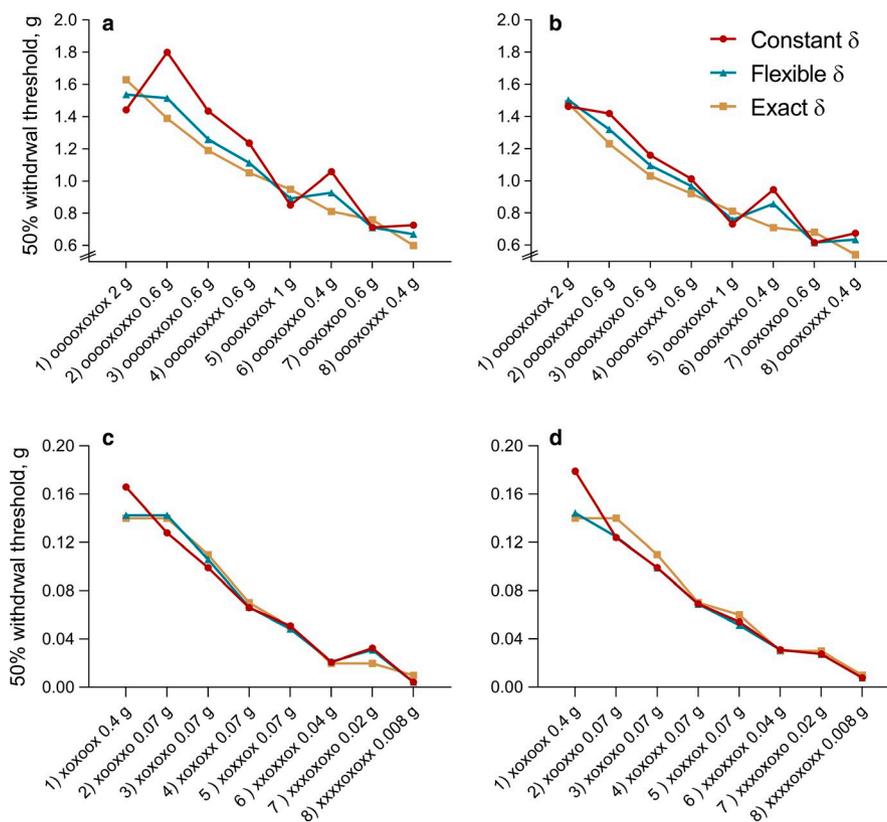


FIGURE 1 XO response patterns and corresponding 50% withdrawal thresholds by six different calculation methods. X-axis; different response patterns of the mouse von Frey scale placed in descending order from left to right. The ranking is based on an objective assessment of the individual thresholds by applying a frequency count of responses to the various filaments (Table 2). Response patterns (O = no response, X = positive response) are followed by the target force, g. of the last filament applied. Y-axis; calculated 50% thresholds, g. for the corresponding XO patterns within the higher range of the scale (a and b) and lower range of the scale (c and d). Left panels (a and c) represent calculations performed by application of the handle number and right panels (b and d) are calculations based on target force. For all panels, the three different interpretations of δ are applied to Chaplan's formula. As XO patterns are placed in descending order, the slope of the curves is expected to decline from left to right. Note the deviation in patterns 1) and 2) in panel a and patterns 5) and 6) in panel a and b. In panels c and d, patterns 1 and 2 are similar according to the frequency of responses to individual filaments, whereas there should be a decline from 6) to 7)

equation does not have the sensitivity to rank the different reaction patterns in the correct descending order. Ranking is improved for response patterns 1) and 2) when changing X_f from handle number to target force (panel a vs. panel b), while patterns 5) and 6) remain inaccurate. Here, the flexible δ value improves the ranking, but only the use of the exact δ completely resolves the problem.

Looking at the low-range patterns in Figure 1c and d, inaccuracies occur in relation to patterns 1) versus 2) and 6) versus 7). Biologically, patterns 1) and 2) are alike in terms of response to the exact same filaments. This result is not obtained when constant δ is applied. Calculations based on handle number also has difficulties detecting the decline from pattern 6) to 7), unless the exact δ is applied.

From Figure 1, it is clear that the traditional up-down method where δ is constant and handle number is inserted as X_f has its limitations. The calculated threshold does not always follow the biological ranking of XO patterns. As a direct result, 50% withdrawal thresholds are under- or overestimated in several cases. Consequently, the traditional method is not applicable to separate test groups that are only marginally different. This raises concerns, in setups comparing efficiency of different compounds that are both effective or when working with smaller groups of animals.

In a theoretical experiment, the efficacy of treatment X and Y in reversing a state of hyperalgesia was tested (raw data are provided as supplementary material) and 50% thresholds calculated by the six different methods. These data show how refinement of the calculation method can increase the power of a dataset. In the first analysis (Figure 2a and Table 3a) of the theoretical dataset the two treatment groups are compared in a parametric t test. With a group size of 12, only the traditional method (constant δ and handle number is inserted as X_f) fails to show a difference between

treatment X and Y. Notably, significance level is increased the more accurately δ is determined. The use of exact δ allows reduction of the group size to six while detecting the difference between X and Y (not shown). Testing the dataset non-parametrically with a Mann–Whitney test and a group size of 12, all six calculation methods detect the difference between treatment X and Y (not shown). However, when group size is reduced to six, only the application of exact δ results in a significant difference between groups; see Figure 2b and Table 3b.

Based on the condition that our interpretation and ranking of data in Figure 1 is correct, it is clearly illustrated that data quality is improved using the exact δ . In the theoretical experiment, the use of exact δ prevented the occurrence of a false negative conclusion (type II error), as we find that treatment X is in fact different from Y, when applying the exact δ in the calculation of withdrawal thresholds. Another situation that raises concern is a case in which a control group (typically within the upper range of the scale) is compared to a test group (within the low range). Here, the occurrence of a type I error is possible when the control group has a high frequency of response pattern 2) (Figure 1a), leading to a significant overestimation of the mean or median if the traditional method is applied. As an identical overestimation does not occur in the low range, an overestimation of the effect size could lead to a false-positive conclusion.

4 | DISCUSSION

When testing efficacy of anti-hyperalgesic drugs, it is essential that the most appropriate and standardized methodology is used in animal studies before proceeding to

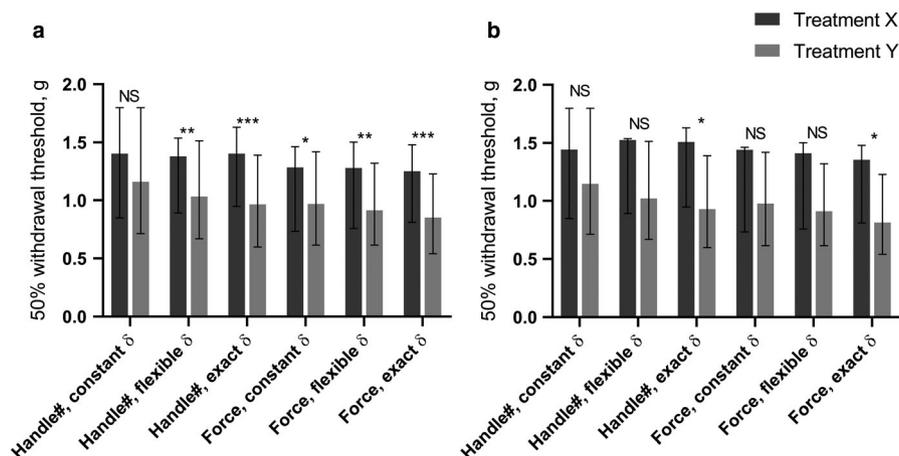


FIGURE 2 Effect of treatments X and Y in a theoretical experiment applying different groups sizes and statistics in a and b, respectively. (a) For X and Y, $n = 12$. Data are shown as mean \pm range and treatment X and Y are compared with a t test for each of the six methods. NS $p \geq .05$, * $p < .05$, ** $p < .01$, *** $p < .001$. (b) For X and Y, $n = 6$. Data are shown as median \pm range and treatment X and Y are compared with a Mann–Whitney test for each of the six methods. NS $p \geq .05$, and * $p < .05$

TABLE 3 Statistics for Figure 2. (a) Multiple *t* tests comparing treatment X and Y. *p* values are uncorrected, *n* = 12. (b) Multiple Mann–Whitney *U* tests comparing treatment X and Y. *p* values are uncorrected, *n* = 6

(a)	<i>p</i> value	Mean of Treatment X	Mean of Treatment Y	Difference	SE of difference	<i>t</i> ratio	df
Handle#, constant δ	0.107124	1.402	1.162	0.2408	0.1434	1.680	22.00
Handle#, flexible δ	0.006542	1.379	1.033	0.3467	0.1154	3.003	22.00
Handle#, exact δ	0.000841	1.403	0.9667	0.4367	0.1130	3.864	22.00
Force, constant δ	0.013358	1.283	0.9712	0.3122	0.1160	2.691	22.00
Force, flexible δ	0.003568	1.280	0.9152	0.3652	0.1119	3.262	22.00
Force, exact δ	0.000957	1.252	0.8517	0.4000	0.1050	3.810	22.00
(b)	<i>p</i> value	Median treatment X	Median treatment Y	Difference	Mann–Whitney U		
Handle#, constant δ	0.16	1.443	1.1147	0.296	9		
Handle#, flexible δ	0.054	1.526	1.021	0.505	6		
Handle#, exact δ	0.041	1.51	0.93	0.580	5		
Force, constant δ	0.054	1.44	0.979	0.460	6		
Force, flexible δ	0.054	1.411	0.9125	0.4985	6		
Force, exact δ	0.041	1.355	0.815	0.540	5		

clinical testing in human experimental pain models or patient populations (Andrews et al., 2016). We show that the Chaplan equation for estimation of 50% withdrawal thresholds needs refinement, especially when applied to mice since the original calculations did not include this range of filaments. With the traditional approach using a constant δ and handle number as X_f , up-down response patterns can be incorrectly ranked resulting in inaccuracies and potentially Type 1 and 2 errors.

Two approaches can be taken to improve the estimations. First, $\log(\text{target force})$ should be inserted as X_f in the Chaplan equation instead of the less accurate handle number; see Table 1. This value better reflects the stimulus applied to the animal and improves the inter filament intervals (only relevant if using traditional definitions of δ); see Table S1. Second, the δ value should be determined more accurately using the flexible or, even better, the exact δ value. This issue has also been addressed by Bradman et al. (2015). Application of the flexible δ can be made by relatively simple changes to the calculation, whereas use of the exact δ requires more advanced statistical insight.

The everyday application of von Frey testing and associated statistics does not allow for scrutiny of the statistical background provided by Dixon (1965, 1980). More often understanding of the methodology is passed from user to user. To reduce inaccuracies, we have developed a free, online, easy-to-use platform for calculation of 50% withdrawal thresholds. Here, users can manually type or upload whole datasets for easy calculation (https://bioapps.shinyapps.io/von_frey_app/). Based on this article, the resource

is built on an R package for individual settings of δ type and range of filaments applied (<https://github.com/MikkeIAStorm/Up.Down.Method>). Crucially, the platform allows the user the possibility to go beyond the Chaplan equation and estimate 50% thresholds using the exact δ value. The improved data quality will increase sensitivity and specificity, potentially making it possible to detect smaller differences between test groups or use fewer animals per group. Notably, this is achieved without changing the data collection process. Hence, users can insert previously collected data to the new algorithm. However, the flexibility of the provided tool allows for any number of filament stimulations beyond the classical Dixon paradigm of four filament presentations following the first XO/OX cross. Modifications to the classical staircase have been proposed by Detloff et al. (2010) suggesting 10 filament application irrespective of reaction pattern, whereas the SUDO method (Bonin, Bories, & De Koninck, 2014) reduces the total number of filament applications to five for all test subjects. Generally, the ideal method should provide a valid estimate of the 50% withdrawal threshold with a minimal number of filament applications as thoroughly discussed by Bradman et al. (2015).

Our work presented here, represents an opportunity to refine the mathematical methodology around von Frey testing while in compliance with the original up-down statistical method developed by Dixon. The refinement provides improved internal and external reproducibility of experimental data and it facilitates transparency between laboratories.

ACKNOWLEDGEMENTS

We wish to thank Dr Tony Yaksh (University of California, San Diego) for his input on interpretation of the traditional up-down method.

CONFLICT OF INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

S.L.C, D.M.K, R.B.H, T.F.H and J.M.G.I initiated the development of the novel algorithm. M.A.S, J.M.G.I and T.F.H developed the algorithm and web app and S.L.C, R.B.H and M.O tested the web app. S.L.C, R.B.H and D.M.K drafted the manuscript. M.O., F.P. and J.O. discussed and revised the manuscript. All authors discussed the results and commented on the manuscript.

REFERENCES

- Andrews, N. A., Latrémolière, A., Basbaum, A. I., Mogil, J. S., Porreca, F., Rice, A. S. C., ... Whiteside, G. (2016). Ensuring transparency and minimization of methodologic bias in preclinical pain research. *Pain*, 157(4), 901–909. <https://doi.org/10.1097/j.pain.0000000000000458>
- Bonin, R. P., Bories, C., & De Koninck, Y. (2014). A simplified up-down method (SUDO) for measuring mechanical nociception in rodents using von Frey filaments. *Molecular Pain*, 10, 26. <https://doi.org/10.1186/1744-8069-10-26>
- Bradman, M. J. G., Ferrini, F., Salio, C., & Merighi, A. (2015). Practical mechanical threshold estimation in rodents using von Frey hairs/Semmes-Weinstein monofilaments: Towards a rational method. *Journal of Neuroscience Methods*, 255, 92–103. <https://doi.org/10.1016/j.jneumeth.2015.08.010>
- Burgos-Vega, C. C., Quigley, L. D., Trevisan dos Santos, G., Yan, F., Asiedu, M., Jacobs, B., ... Dussor, G. (2018). Non-invasive dural stimulation in mice: A novel preclinical model of migraine. *Cephalalgia: An International Journal of Headache*, 39(1), 123–134. <https://doi.org/10.1177/0333102418779557>
- Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M., & Yaksh, T. L. (1994). Quantitative assessment of tactile allodynia in the rat paw. *Journal of Neuroscience Methods*, 53(1), 55–63. [https://doi.org/10.1016/0165-0270\(94\)90144-9](https://doi.org/10.1016/0165-0270(94)90144-9)
- Detloff, M. R., Clark, L. M., Hutchinson, K. J., Kloos, A. D., Fisher, L. C., & Basso, D. M. (2010). Validity of acute and chronic tactile sensory testing after spinal cord injury in rats. *Experimental Neurology*, 225(2), 366–376. <https://doi.org/10.1016/j.expneurol.2010.07.009>
- Dixon, W. J. (1965). The Up-and-Down Method for Small Samples. *Journal of the American Statistical Association*, 60(312), 967–978. <https://doi.org/10.1080/01621459.1965.10480843>
- Dixon, W. J. (1980). Efficient Analysis of Experimental Observations. *Annual Review of Pharmacology and Toxicology*, 20(1), 441–462. <https://doi.org/10.1146/annurev.pa.20.040180.002301>
- Gonzalez-Cano, R., Boivin, B., Bullock, D., Cornelissen, L., Andrews, N., & Costigan, M. (2018). Up-down reader: An Open Source Program for Efficiently Processing 50% von Frey Thresholds. *Frontiers in Pharmacology*, 9, <https://doi.org/10.3389/fphar.2018.00433>
- Milligan, E. D., Mehmert, K. K., Hinde, J. L., Harvey, L. O., Martin, D., Tracey, K. J., ... Watkins, L. R. (2000). Thermal hyperalgesia and mechanical allodynia produced by intrathecal administration of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein, gp120. *Brain Research*, 861(1), 105–116. [https://doi.org/10.1016/S0006-8993\(00\)02050-3](https://doi.org/10.1016/S0006-8993(00)02050-3)
- Mills, C., LeBlond, D., Joshi, S., Zhu, C., Hsieh, G., Jacobson, P., ... Decker, M. (2012). Estimating efficacy and drug ED50's using von Frey thresholds: Impact of weber's law and log transformation. *The Journal of Pain: Official Journal of the American Pain Society*, 13(6), 519–523. <https://doi.org/10.1016/j.jpain.2012.02.009>
- Werner, M. U., Rotbøll-Nielsen, P., & Ellehuus-Hilmersson, C. (2011). Humidity affects the performance of von Frey monofilaments. *Acta Anaesthesiologica Scandinavica*, 55(5), 577–582. <https://doi.org/10.1111/j.1399-6576.2011.02426.x>

SUPPORTING INFORMATION

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

How to cite this article: Christensen SL, Hansen RB, Storm MA, et al. Von Frey testing revisited: Provision of an online algorithm for improved accuracy of 50% thresholds. *Eur J Pain*. 2020;24:783–790. <https://doi.org/10.1002/ejp.1528>