Fitting Psychometric Functions Using a Fixed-Slope Parameter: An Advanced Alternative for Estimating Odor Thresholds With Data Generated by ASTM E679

Mei Peng1,2, Sara R. Jaeger2 and Michael J. Hautus1

1School of Psychology, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand and 2The New Zealand Institute for Plant & Food Research Limited, Mt Albert Research Centre, Private Bag 92169, Auckland, New Zealand

Correspondence to be sent to: Michael J. Hautus, School of Psychology, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. e-mail: m.hautus@auckland.ac.nz

Accepted November 25, 2013

Abstract

Psychometric functions are predominately used for estimating detection thresholds in vision and audition. However, the requirement of large data quantities for fitting psychometric functions (>30 replications) reduces their suitability in olfactory studies because olfactory response data are often limited (<4 replications) due to the susceptibility of human olfactory receptors to fatigue and adaptation. This article introduces a new method for fitting individual-judge psychometric functions to olfactory data obtained using the current standard protocol—American Society for Testing and Materials (ASTM) E679. The slope parameter of the individual-judge psychometric function is fixed to be the same as that of the group function; the same-shaped symmetrical sigmoid function is fitted only using the intercept. This study evaluated the proposed method by comparing it with 2 available methods. Comparison to conventional psychometric functions (fitted slope and intercept) indicated that the assumption of a fixed slope did not compromise precision of the threshold estimates. No systematic difference was obtained between the proposed method and the ASTM method in terms of group threshold estimates or threshold distributions, but there were changes in the rank, by threshold, of judges in the group. Overall, the fixed-slope psychometric function is recommended for obtaining relatively reliable individual threshold estimates when the quantity of data is limited.

Key words: 3AFC, ASTM E679, method of limits, olfactory, psychometric function, threshold

Introduction

In modern psychophysics, estimation of sensory detection thresholds relies on fitting psychometric functions. Broadly defined, a psychometric function describes the dynamic relationship between the stimulus intensity and some performance measure of detection or discrimination (Harvey 1986). The psychometric function increases gradually, primarily as a result of random variation in the sensory system (Green and Swets 1966; Macmillan and Creelman 2005). Just as the unit of stimulus intensity varies with type of stimulus, the performance measure also changes depending on the methodology by which the measure was obtained. In the case of odor detection, the psychometric function conforms to a symmetrical sigmoid shape when the performance is measured by proportion correct ($p(c)$) and the odorant concentration is expressed in logarithmic units. This symmetrical form makes the psychometric function more amenable to modeling.

There are a variety of mathematical models available for defining the psychometric function, the most common ones being Gaussian, logistic, Gumbel, and Weibull (Harvey 1986; Treutwein and Strasburger 1999; Wichmann and Hill 2001). The choice of mathematical model can be made on either theoretical or practical grounds (Harvey 1986). The current work used the Gaussian distribution to define the probability of a correct response as a function of $\log_{10}$ odorant concentration because of its consistency with other standard theories and models and its predominant use in the literature (Green and Swets 1966; Wichmann and Hill 2001). The Gaussian psychometric function can be modeled as $p(c) = \gamma + (1 - \gamma)\Phi(x, \mu, s)$, where $x$ is the $\log_{10}$ odorant concentration and $\Phi$ is the cumulative density function of the normal distribution.
concentration; μ is the mean of the Gaussian distribution, equal to the midpoint of the function on the abscissa (often referred to as the threshold or midpoint threshold); s is standard deviation of the Gaussian distribution, often referred to as the “slope” of the psychometric function (note that as s decreases, the psychometric function gets steeper); and γ is the value of p(c) that is obtained by chance alone (the reciprocal of the number of choices in a multiple forced choice task). This gives the psychometric function 2 asymptotes: γ and 1.

Fitting a psychometric function typically involves repeated sampling of a judge’s performance at several stimulus concentrations, so that p(c) can be calculated for each concentration. Previous studies, using auditory testing, recommended at least 200 test replications at each intensity level to obtain a stable psychometric function with precise estimates for the slope and intercept (Kontsevich and Tyler 1999; Leek et al. 1992). This number reduces significantly if a simple estimation of the threshold is required; in this situation, 30 test replications have been shown to be adequate (Macmillan and Creelman 2005).

Complications arise when incorporating psychometric functions into odor threshold estimation. The primary difficulty lies in the high susceptibility of human olfactory receptors to fatigue and adaptation (Cheesman and Mayne 1955; Ekman et al. 1967). As performance rapidly declines as a result of these factors, experimenters need to carefully consider the quantity and frequency of exposure of judges to odorant samples. Additionally, the time- and resource-intensive nature of conducting olfactory research frequently makes collection of an extensive number of test replications a difficult task.

American Society for Testing and Materials (ASTM) International Standard Practice E679 outlines an efficient data collection method that has served as the standard protocol for odor threshold estimation in sensory research (Lawless and Heymann 1997; Meilgaard et al. 2007). It prescribes the use of a classical psychophysical protocol—the method of limits with an ascending concentration series using 3-alternative forced-choice (3AFC) judgments at each concentration step—to measure odor thresholds for an individual judge. To elaborate, a test session consists of a series of 3AFC presentations, each containing 1 target sample and 2 blank samples. The concentration of the target sample follows a geometric increasing progression at each step in the series. The judge’s task in each 3AFC trial is to identify the target sample. A choice is forced when the judge is unable to detect the target sample using available sensory evidence.

ASTM E679 also includes a threshold calculation heuristic based on evaluating the correctness of the judge’s responses. A judge’s best-estimate threshold (BET) is indicated by the final response reversal from incorrect to correct identification in the concentration series (this is sometimes called the Last Reversal stopping rule; see Peng et al. 2012) and the BET is derived by taking the geometric mean of the concentration at which the last incorrect response occurs and the concentration of the next step higher. When multiple test replications are being administered, a judge’s BET is the geometric mean across all replications. The group BET is the geometric mean of the BETs of all judges. The BETs are often converted to logarithmic (log10) units. This practice transforms the BETs to a scale on which successive concentrations under test are equally spaced and also has the practical effect of normalizing the threshold distribution.

Although the BETs obtained using the ASTM heuristic do not deviate drastically from those obtained with psychometric function-based approaches (Martineau et al. 1995; Robinson et al. 2005; Kolpin and Shellhammer 2009; Lawless 2010; Peng et al. 2012), sensory researchers have never abandoned the idea of developing an alternative with a more solid theoretical basis for odor threshold estimation. Previous studies suggest that there are 2 routes in the development of more sophisticated methods for odor threshold estimation. One is to follow the conventional approach of fitting psychometric functions, collecting a relatively large number of test replications (<30 in most studies) and then fitting the function at the individual level. Examples are the method prescribed by ASTM E1432 and the method outlined by Cometto-Muñiz and associates (Cometto-Muñiz et al. 2002, 2008; Cometto-Muñiz and Abraham 2008, 2009, 2010). The data collection process required by this method is laborious, but advocates believe it is better to undertake this laborious process than to employ an arbitrary heuristic that gives a threshold estimate of unknown nature (Morrison 1982). However, purists might argue that the number of test replications, usually <30, is inadequate for deriving an accurate slope estimate of a psychometric function. Therefore, this method does not fully exploit the benefits that would be expected from using a psychometric function because an imprecise slope parameter potentially reduces stability of the psychometric function and consequently compromises the accuracy of the threshold estimates. Researchers looking to establish odor thresholds need to find the right balance between collecting a large number of test replications from a small number of judges and collecting a smaller number of test replications from a larger group. In the case of fitting conventional psychometric functions, the small sample size limits the estimation of an odor threshold distribution across the population. The combination of these factors has hindered the adoption of the conventional psychometric function in odor threshold estimation.

An alternative approach is to explore a method for fitting psychometric functions with aggregate data. This compensates for a small amount of data from individual judges by combining the responses from all judges. Lawless (2010) and Peng et al. (2012) demonstrated the use of this approach. Both studies provide methods on how to calculate p(c) values across all judges and then fit one group function to illustrate detection performance as a whole. This particular method is expected to have a high applicability in odor
threshold studies because the sample size is usually large enough to produce a reasonable slope measure for the group psychometric function. A downside of this method is that its outcome is a BET for the group only. The inability of the method to obtain BETs for individual judges discourages some researchers from employing this method, especially when the threshold distribution (which requires BETs from individual judges) is of particular interest. An ideal method should therefore be able to precisely determine the threshold estimate for individual judges.

Ideally, sensory researchers would have available a theoretically sound method for odor threshold estimation with a practically viable number of replications and an acceptable computational complexity. This article proposes such a method—fitting psychometric functions with fixed slope to the data of individual judges. This assumes the slope of the psychometric function for different judges is the same for a given odorant, thus the same symmetrical sigmoid function is fitted to the data from different judges by shifting the function laterally to coincide best with the data. Justification for the fixed slope hinges on the fact that previous studies in other senses have demonstrated little difference in the slope of psychometric functions across judges. For example, Harvey (1986) observed that psychometric functions predominantly differ in their locations along the abscissa during visual testing, but not in their slopes. Also, in the rare example of fitting psychometric functions to olfactory data, Stevens et al. (1988) found that the psychometric functions fitted to data from individual judges for a given odorant were highly uniform in the slope parameter. The challenge for this method is to determine what the fixed-slope value should be. Our proposal is to use the slope obtained by fitting the aggregate psychometric function of the group. Implementing this method involves fitting the full psychometric function (both intercept and slope) at the group level and then fitting individual psychometric functions to the data from each judge using the slope parameter that had been estimated for the group function for all judges. Thus, only the intercept is fitted for the psychometric function of each judge.

The most apparent advantage of this approach is that it can be applied to data sets typically obtained in odor threshold testing. Without needing to estimate the slope parameter, a typical data set generated by ASTM E679 is assumed sufficient for fitting psychometric functions at the individual-judge level. In comparison to those psychometric functions fitted to a limited number of test replications that are not able to give a reliable slope measure, using a fixed-slope value reduces the variability of the BET, thus increasing precision of the threshold estimate.

In the reported study, this proposed method—fitting a psychometric function with a fixed-slope parameter—was tested with a data set obtained with ASTM E679. The thresholds estimated using the proposed method were compared with those from the current standard heuristic prescribed in ASTM E679 and with those from conventionally fitted psychometric functions. An additional study was undertaken to explore factors that could potentially affect the slope parameter of aggregate psychometric functions. This information was expected to help refine the method of fitting a fixed-slope psychometric function. If this new method survives scrutiny in these empirical investigations, it may further advance methodologies for odor threshold estimation by aligning techniques with those prevalent in other sensory modalities.

Study 1

Judges

The threshold testing took place at the Sensory Laboratory, The New Zealand Institute for Plant & Food Research Ltd. The test involved a total of 100 judges of Caucasian ancestry (53 females), who were recruited from the general population of Auckland, New Zealand. Ages ranged from 19 to 50 years (Mean = 33.25, standard deviation = 9.79). The ethnicity requirement served to reduce the effect of differences due to any genetic influences on human olfactory sensitivity (Keller et al. 2007). None of the judges had previously served in experiments on odor sensitivity and none self-reported olfactory dysfunction. They were divided into 10 cohorts of 10 judges; 1 cohort was tested at a time. Monetary compensation was given at the end of each session. This study was approved by the Northern X Regional Ethics Committee (NTX/08/11/111). Written informed consent was obtained from each judge. This study complies with the Declaration of Helsinki for Medical Research involving Human Subjects.

Samples

The threshold testing was performed using 10 odorants (coded as F01–F10). Chemical information of each odorant can be found in Table 1. Filtered water served as solvent. A solution of each odorant was diluted to 8 discrete concentration steps with a constant dilution factor per step throughout the scale, to enable a 3AFC method of limits. Details of the dilution protocol used are given in Jaeger et al. (2010).

The concentration progressions for each odorant solution were made following the instructions in ASTM E679 as closely as possible (ASTM, 2004b). The dilution factor for some of the series used might be considered to exceed the recommended value in ASTM E679, which is between 2 and 3. Modifications of the dilution factors were attributed to the relatively large sample size employed in this study. With the number of concentration steps being limited to a maximum of 8, the dilution factor must increase to ensure the concentration series for the particular odorant can encompass the full range of threshold estimates for all judges.
Table 1 Chemical information for the tested odorants

<table>
<thead>
<tr>
<th>Odorant code</th>
<th>Name</th>
<th>Purity</th>
<th>CAS number</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>1,8-cineole</td>
<td>99%</td>
<td>470-82-6</td>
</tr>
<tr>
<td>F02</td>
<td>2-heptanone</td>
<td>99%</td>
<td>110-43-0</td>
</tr>
<tr>
<td>F03</td>
<td>Hircinoic acid</td>
<td>&gt;98%</td>
<td>54947-74-9</td>
</tr>
<tr>
<td>F04</td>
<td>cis-3-hexen-1-ol</td>
<td>98%</td>
<td>928-96-1</td>
</tr>
<tr>
<td>F05</td>
<td>Dipropyl disulfide</td>
<td>98%</td>
<td>629-19-6</td>
</tr>
<tr>
<td>F06</td>
<td>Isobutyraldehyde</td>
<td>99.5%</td>
<td>78-84-2</td>
</tr>
<tr>
<td>F07</td>
<td>Isovaleric acid</td>
<td>98.50%</td>
<td>503-74-2</td>
</tr>
<tr>
<td>F08</td>
<td>Vanillin</td>
<td>99%</td>
<td>121-22-5</td>
</tr>
<tr>
<td>F09</td>
<td>β-damascenone</td>
<td>1.1–1.3 wt. % in 190 proof ethanol</td>
<td>23696-85-7</td>
</tr>
<tr>
<td>F10</td>
<td>β-ionone</td>
<td>96%</td>
<td>79-77-6</td>
</tr>
</tbody>
</table>

The concentration range for each odorant was initially established with reference to the range of thresholds reported in the literature. Results from a pilot test, involving 10 people, corroborated the selected ranges. After commencement of the main study, the concentration series were continually modified on a cohort basis by adjusting the dilution factor and/or the starting concentration step. These modifications were based on judges’ responses to the concentration scale; they were to ensure the inclusion of most judges’ threshold values. Only the most updated concentration series were used for the testing of the following cohorts; the concentration series were proven sufficient upon the completion of the first 3 cohorts.

The odorant samples for each test were prepared by placing 10 mL aliquots of target solution of each concentration into wine glasses and immediately covering with glass lids. Another set of 8 pairs of identical glasses, each containing 10 mL of solvent (filtered water), were made for each judge as the blank samples to accompany the presentation of each target sample. All samples were prepared ~1 h in advance of threshold measurement, to allow equilibration of the headspace, and kept at 22°C.

Procedure

Each testing session consisted of an instruction and 2 threshold tests, with a 20-min break in between. The 2 threshold tests were for 2 different odorants. The judges were required to complete all 8 of the 3AFC trials for each threshold test. The order of presentation of samples within a trial was randomized throughout the test. To mitigate smelling fatigue and adaptation, a 75-s break was enforced between trials. A total of 4 test replications for each odorant were collected from each judge by the completion of the study. The test was carried out in the same manner as in Jaeger et al. (2010).

Data analyses

Testing employed the automated completion of all 8 sets of 3AFC trials via Compusense five 4.2, which for each judge recorded the concentration values and whether the response was correct or not. These data were analyzed in 2 ways. The first approach was as specified in ASTM E679 (see Introduction), noting that when the ASTM method is used, a hypothetical concentration step is required to obtain a threshold estimate should the presented concentration range fail to capture the judge’s response reversal (as explained in the appendices of ASTM E679). This method allows for the construction of a threshold distribution within the group for each odorant.

The other method of analyzing these data was to fit the fixed-slope psychometric function using the group slope estimate. This method entails fitting the psychometric function at the group level prior to fitting the functions at the individual-judge level. To fit the group function, \( p(c) \) was calculated at each \( \log_{10} \) concentration used (\( \log_{10} \) concentrations that were identical to 2 decimal places were collapsed). It should be noted that a small group of concentration series were used in <6% of testing for a particular compound and led to data points on the psychometric functions based on <20 observations. These points typically showed large variability and were omitted from the analysis. Best-fitting Gaussian psychometric functions, corrected for guessing, were then fitted to the group data using maximum-likelihood estimation (McKee et al. 1985). The best-fitting function gives 2 parameters—intercept (\( \mu \)) and slope (\( s \)). The threshold was taken as the estimated value of \( \mu \), which also corresponds to the midpoint (\( p(c) = 0.667 \)) of the psychometric function. The standard errors (SE) of \( \mu \) and \( s \) were obtained from the variance/covariance matrix (see Press et al., 1992, p. 972). The goodness-of-fit of the functions was measured using the statistic \( G^2 \), which is asymptotically distributed as chi-square (Wickens 2002). The degrees of freedom for \( G^2 \) are the number of data points minus the number of parameters to which the model is fitted (2 with the conventional function and 1 with the fixed-slope function). This statistic can be interpreted as the probability that, given the fitted function is correct, the observed data, or more divergent data, could arise from it. Models were assumed inadequate at \( P < 0.001 \) (Bentler and Bonett 1980).

The obtained slope parameter was then used to fit fixed-slope psychometric functions separately for each judge. The intercept parameter that provided the best fit to the data was obtained using maximum-likelihood estimation. The BET for each judge was taken as the intercept parameter of each function, the so called midpoint threshold (at \( p(c) = 0.667 \)). The fixed-slope group BET was taken as the mean across the individual-judge BETs.

Individual thresholds were also estimated by applying the standard ASTM E679 heuristic using the Last Reversal stopping rule. The ASTM group BET was taken as the geometric
mean across individual thresholds. Thresholds estimated with the 2 different methods (fixed-slope psychometric function and ASTM E679 heuristic) were compared at the group level and at the individual level, where particular emphasis was placed on threshold distributions and the judge’s threshold ranking in a group.

Results

The group psychometric functions (fitted with both intercept and slope) had satisfactory goodness-of-fit for all of the 10 odorants ($P > 0.001$). The slope estimate of each function, provided in Table 2, was used in constructing each judge’s fixed-slope psychometric function. The group BET obtained with this method is the mean of the individual-judge threshold estimates. These group BETs were compared with those estimated by the ASTM E679 method, as presented in Table 2, by a paired sample t-test. No significant difference was observed although there was a general trend that the fixed-slope psychometric functions produced higher estimates than those produced by the ASTM method, except for F01 (1,8-cineole) and F06 (isobutyraldehyde).

These 2 methods—fixed-slope psychometric function and ASTM E679 heuristic—were also compared in terms of the distribution of estimated thresholds for the study judges. In general, there was no systematic difference between the threshold distributions estimated by each method. The psychometric functions tended to predict a wider range of sensitivity than that predicted by the ASTM method for most odorants. This was the most evident with F10 ($\beta$-ionone), where 23% of thresholds produced from the psychometric functions exceeded the high end of the threshold range estimated by the ASTM method.

Figure 1 (upper panels) presents the threshold distributions obtained for selected odorants. These 4 odorants were chosen because examples of their threshold distributions are in the literature: F01—1,8-cineole (Pelosi and Pisanelli 1981; Lawless et al. 1995); F06—isobutyraldehyde (Amoore 1976); F09—$\beta$-damascenone (Plotto et al. 2006); and F10—$\beta$-ionone (Plotto et al. 2006). For F06, F09, and F10, both threshold estimation methods replicated the shape of the previously published distributions. Specifically, both methods found a slightly bimodal distribution for F06, a positive skewed distribution for F09 and a bimodal distribution for F10. However, the distributions of F01 thresholds differed between methods; the ASTM method’s distribution was positively skewed, whereas the distribution from the fixed-slope psychometric function showed a tendency toward being bimodal. The distribution obtained with the fixed-slope psychometric function more closely resembles those obtained by Lawless et al. (1995) and Pelosi and Pisanelli (1981).

The next level of comparison between the 2 methods was on the ranking of each judge’s threshold within each threshold distribution. The lower panels of Figure 1 illustrate, for the 4 selected odorants, how the threshold ranking was organized when the threshold measures were obtained with the ASTM heuristic and the fixed-slope psychometric function. On average, 29.4% of judges’ threshold ranking changed by >10 places, with the lowest being 15% for F04 (cis-3-hexen-1-ol) and the highest being 37% for F09 (isovaleric acid) and F08 (vanillin). The largest change in an individual judge’s threshold ranking occurred for F07 when a judge was ranked 30th in 100 judges under the ASTM method but moved to 92nd under the fixed-slope psychometric function method.

According to these ranking results, the choice of method (ASTM E679 or fixed-slope psychometric function) seems to have a considerable effect on the relative position of each judge's threshold.
judge in the threshold distribution. However, group BETs derived from the fixed-slope psychometric function (as the average of individual thresholds) did not differ significantly from the group BETs from the ASTM E679 heuristic. This indicates that the differences in individual thresholds resulted in close to a zero mean difference; an increase in some judges' threshold measures was balanced by a reduction in threshold measures for other judges. This explains the absence of a systematic difference in the distribution shape across the 2 methods.

Study 2

As already noted, the conventional way of fitting psychometric functions requires an extensive amount of data from individual judges. Study 2 focused on a small group of judges each undertaking many replications, instead of a large group of judges undertaking few replications as in Study 1. This provided a data set that enabled fitting of both the conventional and fixed-slope psychometric functions, allowing comparisons between thresholds estimated with these methods.

Methods

Three staff of the Sensory Laboratory (2 females and 1 male, mean age = 33 years) participated in the threshold testing for F04 (cis-3-hexen-1-ol). The testing procedure was the same as outlined in Study 1. All subjects were familiar with the current data collection method. Each judge undertook 10 testing sessions over a course of 3 weeks. Upon completion of the study, a total of 30 test replications, comprising 240 3AFC presentations, were collected from each judge. The concentration series were modified in the first few sessions, so that they were tailored to each judge. This resulted in each concentration being presented a different number of times; fitting the functions took into account the different statistical importance that each concentration carried.

Results

This data set was first utilized to fit a conventional psychometric function for each judge (fitting slope and intercept). The estimated parameters of functions for each judge are given in Table 3. The best-fitting group function had an intercept of 1.419 (SE = 0.146) and a slope of 0.751 (SE = 0.171).

Figure 2 displays the psychometric functions fitted to the data from each judge and to the group data. In the plots, each data point represents a $p(c)$ value at a specific concentration. The $G^2$ statistics, indicating that all 3 functions have adequate goodness-of-fit, are also given in Figure 2.

Next the fixed-slope functions were fitted with the slope of the group psychometric function ($s = 0.751$; for intercept parameters refer to Table 3). The fixed-slope functions also show good conformity to the data (Judge 1: $G^2(8) = 12.63$, $p = 0.12$).
Table 3  Summary of parameter estimates (slope and intercept) of the conventional and the fixed-slope psychometric functions for 3 judges in Study 2

<table>
<thead>
<tr>
<th>Judge</th>
<th>Conventional psychometric function</th>
<th>Fixed-slope psychometric function (group slope = 0.751)</th>
<th>Comparison of BETs between methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>SE</td>
<td>Z-test</td>
</tr>
<tr>
<td>1</td>
<td>0.968</td>
<td>0.248</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>0.398</td>
<td>0.293</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>0.600</td>
<td>0.180</td>
<td>A</td>
</tr>
</tbody>
</table>

The differences between judges or methods were assessed by Z-tests. Results from the comparisons of BETs across methods are expressed in z-scores and P values. Results from the between-judge comparisons are given in the columns labeled Z-test. Different letters indicate that the associated values (either BET or slope) are significantly different (P < 0.05).

Figure 2  Best-fitting psychometric functions for F04 (cis-3-hexen-1-ol) using data collected in Study 2. The first 3 panels present a psychometric function for each judge and the fourth panel presents the group psychometric function using aggregated data. The size of a data point is proportional to the number of times that particular concentration was used; examples of data point size given in the top-left panel apply to the individual-judge functions, whereas those given in the bottom-right panel apply to the group data. Droplines indicate the best-estimate threshold for each judge at \( p(c) = 2/3 \). Goodness-of-fit for each function is indicated by the \( G^2 \) statistics and associated P values.
Several analyses were performed on the data shown in Table 3. Of primary importance, Z-tests showed that the slope parameter of the conventional psychometric function did not vary significantly across the judges. This finding lends support to the hypothesis of invariant slope for the psychometric functions of different judges.

For each judge, Z-tests did not reveal significant differences in the BETs between the 2 methods. Also, the rank order of judges in terms of the BET remained unchanged for both methods. However, the BET shifted between methods, so that the BET of Judge 2 was significantly different from that of Judge 1 when using conventional psychometric functions but was significantly different from Judge 3 when the fixed-slope psychometric functions were fitted.

Figure 3 presents a direct visual comparison between the psychometric functions that were fitted with the 2 different methods. The position of Judge 2 shifted from being closer to Judge 3 with the conventional psychometric functions to being closer to Judge 1 with the fixed-slope psychometric functions. Such changes in proximity between the BETs obtained by judges relate to the changes in sensitivity ranking observed in Study 1. These findings demonstrate changes in threshold distribution in response to the choice of method.

It is worth noting that if threshold estimates are derived from the conventional psychometric function, sensitivity rankings will vary depending on the value of \( p(c) \) chosen to define the threshold. In contrast, the rankings predicted by the fixed-slope psychometric function are robust to changes in the value of \( p(c) \). As an example, Judge 2 is ranked as the lowest, middle, or highest BET of the group, depending on whether a low, middle, or high value of \( p(c) \) defines the threshold. In the present case, these changes in ranking are not apparent unless an extreme \( p(c) \) is chosen, but changes in ranking may be expected with more judges even if the fitted slopes do not differ significantly. With the fixed-slope function, a judge always has the same ranking in the threshold distribution no matter what definition of threshold is used.

Study 3

In Study 3, analyses were performed to gain further insight into the fixed-slope method for fitting psychometric functions to the data from individual judges. Specific objectives included testing the validity of the method and uncovering factors that affected the slope of the psychometric function.

Methods

Differences in the slopes of psychometric functions for groups of judges with different average levels of sensitivity were evaluated. The data set described in Study 1 allowed such an analysis. The 100 judges were divided into quartiles based on their BETs obtained with the ASTM method. Each quartile consisted of 25 judges, each of whom had 4 test replications. This gave approximately 100 replications at each concentration. To recapitulate the procedure, not all concentration series were used the same number of times.
The group psychometric function for each quartile and each odorant were examined and compared with each other.

### Results and discussion

Table 4 presents a summary of the parameter estimates for the best-fitting psychometric functions for the most (Q1) and the least (Q4) sensitive quartiles. Differences in the slopes across quartiles were assessed by Z-tests. Significant differences in the slope were observed for half of the odorants. This finding suggests a linkage between the intercept and the slope of the psychometric functions. Furthermore, a significant positive correlation was observed between the SE of the BET and the slope (see Table 4). This finding prompted further investigation of the relationship between these 2 variables.

Figure 4 plots the differences between Q1 and Q4 in terms of the slope as a function of the differences between these 2 quartiles in terms of the SE of the BETs. The regression line \( r = 0.747, P = 0.013 \) indicates that an increase in the difference between the SE of the BETs from Q1 to Q4 is associated with an increase in the difference between the slopes. Furthermore, the y-intercept in Figure 4 is at about \(-0.1\), suggesting that no change in the SE of BETs between Q1 and Q4 corresponds to a minimal change in the slopes between Q1 and Q4.

These 2 difference variables may be related because they were estimated using the same fitting process. To alleviate the possibility that the observed relationship was a consequence of this association, the analyses were repeated using the same slope estimates, but the SE of the BETs were derived from the ASTM method. These SEs were obtained independently of the psychometric functions. A significant correlation was also obtained between the differences in the slopes (psychometric functions) and the differences in the SE of the BETs derived from the ASTM method \( r = 0.825, P = 0.003 \).

When a psychometric function is fitted to group data, it is in fact being fitted to the average of multiple psychometric functions from different judges. The BET estimated by the group function, equivalent to its intercept, primarily determines the location of the function on the concentration axis. Thus, the SE of the BET in part quantifies the variability of the BETs for individual-judge functions. This variability could be viewed as providing an indication of how widely these individual-judge psychometric functions are distributed along the axis. Thus, a small SE of BET implies the individual-judge functions are proximate to each other and conversely, a large SE of BET implies they are more widely distributed.

Figure 5 illustrates how the distribution of BETs directly influences the slope. Each panel consists of the psychometric functions for 3 judges (solid lines), having identical slopes. The 3 functions in the left panel are distributed more widely than those in the right panel. When the functions of the 3 judges are averaged into a group function (dotted line), the shape of the group function is different from those from which it was constructed. The group function in the left panel appears flatter (a higher slope estimate), as opposed to the function in the right panel. This illustrates that the group function always differs from the individual-judge psychometric functions, and the difference is larger if the individual-judge functions are more variable.

### Table 4

Summary of the parameter estimates for the psychometric functions fitted to the most sensitive (Q1) and least sensitive quartile (Q4) of judges in Study 3

<table>
<thead>
<tr>
<th>Odorant</th>
<th>Q1 BET</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>Q4 BET</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>-0.354</td>
<td>0.096</td>
<td>0.664</td>
<td>0.085</td>
<td>1.935</td>
<td>0.179</td>
<td>1.004</td>
<td>0.258</td>
</tr>
<tr>
<td>F02</td>
<td>1.342</td>
<td>0.076</td>
<td>0.574</td>
<td>0.071</td>
<td>2.430</td>
<td>0.041</td>
<td>0.335</td>
<td>0.035</td>
</tr>
<tr>
<td>F03</td>
<td>2.566</td>
<td>0.061</td>
<td>0.464</td>
<td>0.059</td>
<td>3.612</td>
<td>0.081</td>
<td>0.722</td>
<td>0.137</td>
</tr>
<tr>
<td>F04</td>
<td>1.453</td>
<td>0.081</td>
<td>0.548</td>
<td>0.077</td>
<td>2.832</td>
<td>0.081</td>
<td>0.753</td>
<td>0.074</td>
</tr>
<tr>
<td>F05</td>
<td>-0.103</td>
<td>0.043</td>
<td>0.334</td>
<td>0.051</td>
<td>1.139</td>
<td>0.053</td>
<td>0.611</td>
<td>0.071</td>
</tr>
<tr>
<td>F06</td>
<td>-0.187</td>
<td>0.099</td>
<td>0.752</td>
<td>0.108</td>
<td>1.042</td>
<td>0.134</td>
<td>1.000</td>
<td>0.132</td>
</tr>
<tr>
<td>F07</td>
<td>2.919</td>
<td>0.081</td>
<td>0.573</td>
<td>0.085</td>
<td>3.833</td>
<td>0.059</td>
<td>0.491</td>
<td>0.065</td>
</tr>
<tr>
<td>F08</td>
<td>2.559</td>
<td>0.058</td>
<td>0.462</td>
<td>0.066</td>
<td>3.330</td>
<td>0.066</td>
<td>0.607</td>
<td>0.074</td>
</tr>
<tr>
<td>F09</td>
<td>-2.577</td>
<td>0.083</td>
<td>0.584</td>
<td>0.072</td>
<td>-0.775</td>
<td>0.106</td>
<td>1.169</td>
<td>0.115</td>
</tr>
<tr>
<td>F10</td>
<td>-0.286</td>
<td>0.096</td>
<td>0.620</td>
<td>0.095</td>
<td>3.048</td>
<td>0.074</td>
<td>0.436</td>
<td>0.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference in slope (Q1 – Q4)</th>
<th>z-score</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.252</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>3.019</td>
<td>0.001*</td>
<td></td>
</tr>
<tr>
<td>-1.730</td>
<td>0.042*</td>
<td></td>
</tr>
<tr>
<td>-1.920</td>
<td>0.027*</td>
<td></td>
</tr>
<tr>
<td>-3.169</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>-1.454</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>0.766</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>-1.462</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td>-4.312</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>1.598</td>
<td>0.055</td>
<td></td>
</tr>
</tbody>
</table>

Pearson’s correlation coefficients between the SE of the BET and the slope (adjacent shaded columns) are given in the bottom row. The differences in slopes between Q1 and Q4 were assessed by z-tests.

* P < 0.050.
This observation was empirically tested by comparing the slope estimates from each of the quartile psychometric functions to the slope estimate for the overall psychometric functions based on the data from all quartiles and thus the responses from all 100 judges. The expectation was that the overall psychometric function would have a larger slope parameter (i.e., be less steep) than those of the quartile functions because the overall function included a larger range of BETs across the functions of individual judges.

Figure 4  Plot of the difference in the SE of the BETs against the difference in the slope estimates, between the most sensitive (Q1) and the least sensitive quartile (Q4) for Study 3. A linear regression line (dashed line) is fitted to illustrate the relationship between these 2 variables. The dotted lines indicate the intercept and suggest that no change in the SE of the BETs leads to a minimal change in slope.

Figure 5  Illustration of the relationship between the range of the BETs across fixed-slope psychometric functions and the slope of the group psychometric function. Each panel illustrates how a group function (dotted curve) has a different shape (slope) from the set of 3 individual-judge fixed-slope psychometric functions from which it was constructed. The slope of the group function is influenced by the level of variation in the BETs with a larger change in slope with greater variability, illustrated in the left panel compared with the right panel.
Assessments of the function slopes confirmed that each quartile slope estimate was smaller than that for the corresponding overall function (the overall slope estimates are presented in Table 2). A series of Z-tests was performed to assess the differences between the slope estimates for the overall function and the quartile functions for each odorant. Table 5 shows that in 31 out of 40 cases, the slope estimate for the overall function was significantly different from the quartile functions. The shaded sections of Table 5 indicate the nonsignificant comparisons.

**General discussion**

ASTM E679 is one of the most commonly employed procedures for measuring odor thresholds. It prescribes a simple and prompt data collection process combined with a straightforward threshold calculation heuristic. Although thresholds obtained with ASTM E679 have proven adequate, a logical next step would be to incorporate psychometric functions into the analysis of data collected with ASTM E679, so that odor threshold measurement could be considered within an existing psychophysical framework. Previously proposed methods to achieve this include ASTM E1432 (ASTM 2004a), which fits conventional psychometric functions to data from individual judges (Cometto-Muñiz et al. 2002; Cometto-Muñiz and Abraham 2009), and Lawless’ (2010) graphical method for ASTM E679 data that fits psychometric functions at the group level (Peng et al. 2012). A number of recent studies of odor thresholds show an increasing trend toward applying psychometric functions (Cometto-Muñiz and Abraham 2009; Kolpin and Shellhammer 2009; Tempere et al. 2011). However, due to the challenges with collecting large quantities of olfactory response data, the existing methods typically produce either individual BETs for a few judges or a group BET. These methods may be less useful in odor threshold studies when the primary interest lies in threshold distributions across a population. We introduce a proposed new approach to address the above, which we have compared with the standard ASTM method. Our assumption that judges share a constant slope across their separate psychometric functions allows construction of individual-judge psychometric functions with smaller (practically sized) data sets such as those obtained with ASTM E679. With regard to group BETs, there was little difference between those derived from the fixed-slope psychometric function and the ASTM method. This finding is consistent with previous comparative studies that observed a close alignment between the BETs from the ASTM method and from alternative methods based on psychometric functions (Kolpin and Shellhammer 2009; Lawless 2010; Cliff et al. 2011; Peng et al. 2012). The equivalence between these 2 methods in terms of group BETs supports continued use of ASTM E679. However, Lawless (2010) ascertained that ASTM E679 is associated with 2 types of bias that affect its estimated BETs for individual judges: one leads to underestimated BETs and the other leads to overestimated BETs. This assertion challenges the validity of the ASTM method by proposing that a reasonable group estimate from the ASTM method results from a fortuitous cancellation between the 2 types of bias.

This study provided an opportunity to verify Lawless’ (2010) assertion by comparing the BETs for individual judges obtained by the fixed-slope psychometric function and the ASTM method. In this study, comparisons of the threshold distributions between these 2 methods did not show considerable changes. However, it was found that almost one-third

### Table 5 Summary of the Z-tests in Study 3 between the slope estimate of each quartile function (Q1 – Q4 according to sensitivity) and that of the corresponding overall function

<table>
<thead>
<tr>
<th>Odorant</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$z$-score</td>
<td>$P$</td>
<td>$z$-score</td>
<td>$P$</td>
</tr>
<tr>
<td>F01</td>
<td>6.620</td>
<td>&lt;0.001</td>
<td>4.054</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F02</td>
<td>0.897</td>
<td>0.185</td>
<td>1.370</td>
<td>0.085</td>
</tr>
<tr>
<td>F03</td>
<td>3.561</td>
<td>&lt;0.001</td>
<td>3.092</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F04</td>
<td>3.923</td>
<td>&lt;0.001</td>
<td>2.457</td>
<td>0.007</td>
</tr>
<tr>
<td>F05</td>
<td>6.472</td>
<td>&lt;0.001</td>
<td>3.778</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F06</td>
<td>7.035</td>
<td>&lt;0.001</td>
<td>6.466</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F07</td>
<td>0.880</td>
<td>0.212</td>
<td>1.381</td>
<td>0.084</td>
</tr>
<tr>
<td>F08</td>
<td>2.088</td>
<td>0.018</td>
<td>3.309</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F09</td>
<td>8.290</td>
<td>&lt;0.001</td>
<td>6.180</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F10</td>
<td>8.221</td>
<td>&lt;0.001</td>
<td>9.932</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Slope estimates for Q1 and Q4 are presented in Table 4 and those for the entire group in Table 2. The slope estimates for Q2 and Q3 are not presented. The shaded sections contain the comparisons with no significant difference ($P > 0.050$).

Downloaded from https://academic.oup.com/chemse/article-abstract/39/3/229/503441 by guest on 17 November 2018
of the judges, averaged across the odorants, changed their sensitivity rankings considerably (by >10%) when the method by which the thresholds were estimated changed from ASTM E679 to the fixed-slope psychometric function. Thus, although individual thresholds were estimated to be different by the 2 methods, these differences led to no change in the average group BETs. Threshold changes at the individual level, when averaged over individuals, effectively canceled each other out, resulting in the seemingly unchanged group threshold measures and threshold distributions. This finding is considered the first empirical verification for Lawless’ (2010) proposal. We conclude that the individual-judge BETs obtained by the fixed-slope psychometric functions are more reliable than those obtained by Lawless’ method of ASTM E679 is vulnerable to responding errors or correct guesses.

The above discussion does raise the question whether a reliable group threshold measure or threshold distribution necessarily proves the reliability of the ASTM method. The ideal method should not only be able to provide an accurate estimate of the group threshold and a reliable threshold distribution across the population but also be able to correctly rank judges by their BETs. The last requirement has been largely neglected in previous assessments of methods for odor threshold estimation. It perhaps deserves more attention given its relevance in panelist screening, off-flavor detection, identification of specific anosmia, and any other type of study where a judge’s threshold measure is of particular interest. Examples of these kinds of studies include Siegmund and Pöllinger-Zierler (2006), Botezatu and Pickering (2012), and Harwood et al. (2012).

In comparison to the psychometric function fitted in the conventional way, the fixed-slope psychometric function has a few unique advantages besides the theoretical superiority historically associated with use of psychometric functions. First, application of this method does not require a large data set as required for fitting the conventional psychometric function. A common data set collected via ASTM E679, comprising as few as 4 test replications from each judge, would be sufficient for estimating a threshold for each judge. Second, the fixed-slope function establishes relations between the judges’ thresholds in a consistent manner. In other words, the sensitivity ranking is invariant of the performance criterion used to define the “threshold.” This feature could be advantageous as it adds some constancy into establishing a sensory profile for judges. Finally, the fixed-slope psychometric function is likely to be more robust than the conventionally fitted psychometric functions. Imprecise slope estimates would likely introduce additional bias to the threshold estimates, so the fixed-slope psychometric function allows a better estimate of each judge’s threshold than that allowed by the conventionally fitted function given a fixed number of replications. Although the present research uses odor thresholds, we suggest the results are readily applicable to other chemosensory modalities.

Results from Studies 2 and 3 help scope the direction for future research implementing the fixed-slope psychometric function. Study 2 provides some promising evidence that the assumption of invariant slope across judges could have validity. Although the sample size was small, the functions fitted with the fixed-slope psychometric function did not show poorer fits than those fitted in the conventional manner. This tentatively supports the proposed method. Findings from Study 3 challenge the use of group slope for implementing the fixed-slope function. Analyses revealed that the slope parameter of a psychometric function depends on the range of thresholds of judges for which the function accounts. Because a judge’s threshold falls into a small window in the population range, the slope estimated for the group function must be larger than the individual fixed slopes. Further research should investigate whether a more accurate fixed-slope value could be obtained and determine whether or not such an additional refinement of the method makes any practical difference to estimated thresholds.

Funding

The Ministry of Business, Innovation and Employment, New Zealand is gratefully acknowledged for supporting this work (C06X0805).

Acknowledgements

We thank our collaborators at The New Zealand Institute for Plant & Food Research Limited for their input and assistance with collecting and processing data: K. R. Atkinson, L. G. Axten, C. M. Bava, M. K. Beresford, S. L. Cheang, J. Gamble, F. R. Harker, D. Jin, R. D. Newcomb, A. G. Paisley, S. Rouse, N. de Silva and M. W. Wohlers. Special thanks go to Dr Benedicte Pineau and Dr Jeremy McRae for their valuable comments. Professor Åke Hellström contributed valuable feedback and suggestions for improvement during the review process.

Disclosure statement

Dr M.J.H. reports that Ms M.P. and Dr S.R.J. are employed by The New Zealand Institute for Plant & Food Research Limited, and that the research was supported by a grant to The New Zealand Institute for Plant & Food Research Limited from The Ministry of Business, Innovation and Employment, New Zealand.

References


