The Derivation of Optimum Criteria for Use in the Monothermal Caloric Screening Test

Guy Lightfoot,[1] Fiona Barker,[2] Keri Belcher,[3] Vicki Kennedy,[1] Greg Nassar,[4] and Frances Tweedy[5]

Objective: The aim of this study was to determine whether it is possible to predict a normal outcome of the bithermal caloric test by testing at a single temperature and if so, what criteria are most appropriate to use.

Design: A total of 490 patients were considered candidates for the bithermal test and 414 completed the four necessary components, their nystagmus being measured using videonystagmography.

Results: Clinical decision analysis revealed that the cool monothermal test does not provide an adequate combination of sensitivity and specificity for us to recommend its clinical use. However, the warm monothermal test offers a sensitivity of 95% with 29% of patients with normal bithermal results having to undergo the bithermal test (specificity = 71%) if a combination of three criteria are used: a normal bithermal caloric test outcome can be anticipated and testing curtailed after the first temperature if (a) the warm monothermal caloric asymmetry (MCA) is <15% and (b) the two warm results are each >8 degrees per sec (°/sec), and (c) any spontaneous nystagmus is <4 °/sec.

Conclusions: When appropriate criteria are used, the warm monothermal caloric test offers a performance that is acceptable for routine clinical use, sparing a considerable proportion of patients from unnecessary tests at the cool temperature. We believe that the warm/cool monothermal test difference is probably a consequence of the interrelationship between canal paresis and directional preponderance.

(Ear & Hearing 2009;30;54–62)

INTRODUCTION

The standard bithermal caloric test forms an important part of the evaluation of vestibular function in the balance-impaired or dizzy patient. Conventionally, it subjects the patient to a total of four caloric stimuli, warm and cool in each ear. For each, the resulting nystagmus is recorded and the peak nystagmus slow-phase velocity (SPV) can be used as a measure of the sensitivity of that ear to that stimulus. From the four SPV results, two measures may be computed, canal paresis (CP) (also commonly known as unilateral weakness) that is a right/left ear comparison and directional preponderance (DP) that is right/left beating nystagmus comparison. Jongkees (1962) formulae for CP and DP are as follows:

CP is given by:

$$\frac{(\mathrm{WR} + \mathrm{CR}) - (\mathrm{WL} + \mathrm{CL})}{\mathrm{WR} + \mathrm{WL} + \mathrm{CR} + \mathrm{CL}} \times 100\%$$

[1]Department of Clinical Engineering, Royal Liverpool University Hospital, Prescot Street, Liverpool, England; [2]East Berkshire Audiology Unit, King Edward VII Hospital, Windsor, England; [3]Department of Audiology and Hearing Therapy, Royal South Hants Hospital, Brinton's Terrace, Southampton, England; [4]Audiology Department, Withington Community Hospital, Nell Lane, West Didsbury, Manchester, England; and [5]Audiology Department, Manchester Royal Infirmary, Oxford Road, Manchester, England.

DP is given by:

$$\frac{(\mathrm{WR} + \mathrm{CL}) - (\mathrm{WL} + \mathrm{CR})}{\mathrm{WR} + \mathrm{WL} + \mathrm{CR} + \mathrm{CL}} \times 100\%$$

where WR = warm right; WL = warm left; CR = cool right; and CL = cool left.

The British Society of Audiology (BSA, 1999) published a Recommended Procedure for the bithermal caloric test in 1999, which suggested that centers establish their own criteria for identifying abnormal CP and DP. Values between 20 and 25% are most commonly used, corresponding typically to the 95% confidence limit of these measurements in normal subjects. The caloric test is generally well tolerated but in a small percentage of cases, the patient finds the extent of the induced vertigo distressing and very occasionally, the test has to be curtailed. Several studies (Table 2) have sought to establish the concept of a caloric screening test, based on just one stimulus temperature, applied to each ear in turn. Here, if certain criteria that predict a normal bithermal test outcome are satisfied, it is statistically highly unlikely that the results of tests using the second temperature will change the outcome and therefore the second temperature stimulus need not be used. Patients in whom these criteria are not met immediately go on to receive the second temperature and thus complete the bithermal test to define the extent of any CP and/or DP. The caloric screening test therefore carries the promise of saving time and minimizing the patient's exposure to potentially unpleasant stimuli without sacrificing diagnostic accuracy in cases where this is appropriate. We do not intend to suggest that the screening test is performed in addition to or on a separate occasion to the bithermal test; rather, when performing the caloric test, after the second irrigation the two available results are inspected to see whether a normal outcome is already evident and if so, testing is curtailed at that point.

Index of Asymmetry

Several studies have investigated the clinical utility of a monothermal caloric screening test. The monothermal caloric asymmetry (MCA) is calculated from the formula:

$$\mathrm{MCA}\ \% = \frac{(\mathrm{R} - \mathrm{L})}{(\mathrm{R} + \mathrm{L})} \times 100\%$$

where R and L are the peak SPV recorded from the right and left ear tests using the single irrigation temperature. Most studies have used the MCA by selecting an asymmetry criterion arbitrarily or by choosing a confidence interval (e.g., 90%) from the MCA distribution in subjects with no significant CP or DP. To do this is to set a known specificity (the ability of the MCA to correctly predict a normal outcome), with the hope

0196/0202/09/3001-0054/0 • Ear & Hearing • Copyright © 2009 by Lippincott Williams & Wilkins • Printed in the U.S.A.

TABLE 1. Worked examples illustrating a normal subject, a patient with a left-sided canal paresis (CP) only and a patient with a left-sided CP with a directional preponderance (DP) to the right

	Warm right (°/sec)	Warm left (°/sec)	Cool right (°/sec)	Cool left (°/sec)	CP (%)	DP (%)	Warm MCA (%)	Cool MCA (%)
Normal	24.3	25.7	21.9	22.2	1.8	1.2	2.8	0.7
CP only	26.7	11.3	21.4	10.8	**37.0**	6.8	**40.5**	**32.9**
CP and DP	40.6	17.1	15.1	14.3	**27.9**	**26.1**	**40.7**	2.7

CP and DP relate to the results of the bithermal caloric test. MCA is the monothermal caloric asymmetry. Figures in bold indicate values that would be interpreted as abnormal.

that an adequate sensitivity is achieved. We believe this to be the wrong approach. Although a high specificity (low false-positive rate) is desirable, far more important is sensitivity. This is the ability of the MCA to correctly predict when the bithermal caloric test will show an abnormal outcome. A poor sensitivity (a high false-negative rate) would suggest that the MCA is normal (and the bithermal caloric test is therefore not performed) in cases where an abnormal result of the bithermal test may have been obtained. Minimizing such false negatives must be the primary factor governing the choice of MCA criterion, as advised by Enticott et al. (2003).

Other Criteria

In addition to the MCA criterion, one needs to consider the circumstances in which a normal MCA result might appear in the presence of a genuine CP or DP. One such scenario is bilateral hypofunction, so a reasonable additional criterion is to specify a lower absolute limit for the two SPV results. Another is the combined presence of a CP and DP. The relevance of the interaction between CP and DP is complex and is best illustrated by an example.

Table 1 considers three scenarios and suggests typical caloric results. The first is a normal case. The second is a case of left CP but with an insignificant DP. Note that the warm and cool monothermal MCA values are high, and therefore, this sort of case would normally be correctly identified by both monothermal tests. The third example has a left CP, this time with a DP to the right, as is often typical in cases of incompletely compensated left-sided vestibular dysfunction. The DP to the right will often manifest itself as a spontaneous nystagmus to the right (Baloh, et al., 1977; Jacobson, et al., 1997). However, a DP is not always opposite in direction to a CP.

This combination of CP and DP often has the effect of producing two somewhat similar cool stimulus SPV results and two very different warm stimulus SPV results. An abnormality would therefore be clearly evident in the two warm results but may be missed only if the cool results are considered as illustrated in Table 1. It is for this reason that the warm MCA has been more popular than the cool MCA, because it is likely to be associated with fewer false-negative results. A third criterion is, therefore, helpful to further reduce the likelihood of false-negative screening test results: that of any evidence of a potential DP, notably any significant spontaneous nystagmus in the precaloric test condition (CTC), gaze test, or other test in which the presence of nystagmus is an abnormal result. These additional criteria should result in a further reduction of the false-negative rate albeit at the expense of a greater number of false positives. Three criteria are therefore appropriate, and the bithermal caloric test should be completed if any of the following are not satisfied:

- the MCA must be below a certain percentage
- both monothermal SPV results must be greater than a certain value to exclude cases of bilateral hypofunction
- there must be no average spontaneous nystagmus faster than a certain value

Table 2 is a distillation of the literature on this topic. All studies used electronystagmography (ENG) recording of nystagmus, but none made any reference to the steps taken to ensure optimal calibration of the recording system, notably recalibration immediately before individual irrigations. Most studies did not report on the order of the four irrigations. These two limitations combine to cast doubt on the accuracy and validity of these studies, the danger being that uncompensated corneoretinal changes over time, coupled with a fixed test order, could be mistaken for physiological adaptation of the vestibular system (Lightfoot, 2004). From examination of Table 2, it is apparent that poor MCA false-negative rates (i.e., poor sensitivity) are associated with MCA criteria that exceed 20% or do not include subsidiary criteria. This is a logical consequence of choosing an MCA criterion that is similar to or greater (more lax) than the bithermal criterion for CP. Although many studies have suggested that the performance of the warm MCA is superior to that of the cool MCA, the study by Enticott et al. (2003), using an air stimulus, found the opposite. Their choice of a very strict 5% MCA criterion yielded an impressive cool MCA false-negative rate of about 1% (i.e., 99% sensitivity) but with correspondingly poor (78%) false-positive rate that led the authors to conclude that there was little point in adopting it routinely.

The aim of this study was to define the criteria required to safely curtail the test after one temperature when those results suggest that an abnormal bithermal test outcome is highly unlikely. It is not intended that the monothermal test be used as an alternative to the bithermal test. We believe that for a monothermal test to be worthwhile, the false-negative rate must be below 5% (sensitivity >95%) and that the false-positive rate should be below 50% (specificity >50%). Such a poor specificity may seem inappropriate. However, a monothermal positive outcome (be it a true or false positive) merely leads to the performance of the two irrigations at the second temperature, after which the definitive outcome, in the form of the bithermal caloric result, is available. Indeed, any specificity greater than zero constitutes a saving of time and carries no significant penalty providing sensitivity is maintained. This study aims to investigate whether these aims can be realized and if so, establish the necessary MCA criterion values by examination of a large number of bithermal caloric tests from a typical clinical population.

TABLE 2. A summary of findings from the literature on the monothermal caloric test

Study	No. subjects	Stimulus	Definition of CP (%)	Temperature	MCA criterion (%)	Minimum °/sec nystagmus	Other criteria	False neg (%)	False Pos (%)
Enticott et al. (2003)	744	50 °C and 24 °C air, 60 sec 8 L/min	>25	Warm	5	*	*	7.1	67
				Cool	5	*	*	0.8	78
Farid et al. (2003)	97	250 mL in 40 sec water	>20	Warm	30	†	†	39	13
				Cool	27	†	†	33	13
Jacobson et al. (1995)	504	250 mL in 40 sec water	>20	Warm	24.5	11	‡	6.6	4
Keith et al. (1991)	200	250 mL in 40 sec water	>20	Warm	20	None	None§	29§	8.5
				Cool	20	None	None§	36§	5.7
Norre (1987)	272	Water, 30 sec	>20	Warm	10	5	None	5.7	48
				Cool	10	5	None	5.7	45
Jacobson and Means (1985)	30	250 mL in 40 sec water	>22	Warm	29.5	11	‡	0¶	5
Becker (1979)	362	‖	>20	Warm	20	11	None	14	22
				Cool	20	11	None	25	15

Minimum °/sec is a criterion to identify cases of bilateral hypofunction. Both results of a monothermal test must exceed this value.
We define the false-negative rate as the proportion of cases with a significant canal paresis (CP) or directional preponderance (DP) (as identified by the bithermal caloric test, using the criteria given) that were missed by the monothermal screening test used in the study, using all available criteria proposed by the authors. We define the false-positive rate as the proportion of cases with no significant CP or DP (as identified by the bithermal caloric test, using the criteria given) that were falsely failed by the monothermal screening test used in the study, using all available criteria proposed by the authors.
All studies using a water stimulus employed temperatures of 44 °C and 30 °C.
** Multiple criteria were used to identify bilateral hypofunction and these cases, together with any spontaneous nystagmus were excluded (the criterion for spontaneous nystagmus was not specified in the article but was >2 °/sec, and the warm false-positive rate was 67%, established by personal communication).*
† Patients with bilateral hypofunction, hyperfunction, spontaneous nystagmus, or other caloric abnormality were excluded from this study. Screening test criteria were derived from 95% limits.
‡ Abnormal ocular motility, positional, or spontaneous nystagmus (undefined).
§ Most false-negative findings occurred when the patient's CP was in the range 16–30%. The authors suggest that monothermal test would be inappropriate (the full caloric test should be conducted) if there is any spontaneous nystagmus, but they do not state whether any of their subjects had this or any other abnormal finding.
¶ Assessed on only 13 subjects with a significant canal paresis.
‖ Either water (44 and 30 °C, 250 mL in 40 sec) or air (50 and 24 °C, 9 L in 60 sec).
CP, canal paresis; MCA, monothermal caloric asymmetry; False neg, false negative rate (equivalent to 1-sensitivity); False pos, false positive rate (equivalent to 1-specificity).

MATERIALS AND METHODS

Four English National Health Service Hospital audiology departments contributed to this study.

- Center A: Royal Liverpool University Hospital
- Center B: King Edward VII Hospital, Windsor
- Center C: Manchester Royal Infirmary
- Center D: Withington Hospital, Manchester

Subjects

A total of 490 patients were candidates for the bithermal caloric test over the period of December 2006 to October 2007 (162 men, 328 women, age range 10 to 87 yr, mean age 49.9 yr). These patients had a history of vertigo or imbalance, referred for standard diagnostic vestibular function tests as part of their diagnostic evaluation. Tests included tympanometry, saccadic and smooth pursuit tests, gaze deviation, and caloric tests. Other tests (e.g., Dix-Hallpike or static position) were also performed as dictated by the patient's symptom profile.

Procedures

All centers used videonystagmography (VNG) and caloric stimuli meeting the specification detailed in the BSA Recommended Procedure: centers A, C, and D used water (open loop) 44 °C and 30 °C for 30 sec, 500 mL/min, whereas center B used otoscopy-assisted air 50 °C and 24 °C for 60 sec, 8 L/min, parameters that are believed to provide approximately equivalent water and air stimuli. This was to facilitate an examination of any differences in monothermal performance obtained by these stimulation media. Patients were supine with their head raised 30° to bring their lateral semicircular canals into the vertical plane. In each center, tests were conducted or directly supervised by an experienced audiologist for whom caloric testing was an everyday procedure. Particular attention was paid to the quality and efficiency of irrigations and only trivial amounts of cerumen were tolerated. The sequence of caloric stimuli was preceded by a practice run in which the average velocity of any spontaneous nystagmus in the caloric test condition was noted. Patients were required to undergo a mental alerting task during this and subsequent VNG recording of caloric nystagmus.

VNG was used to avoid the confounding effects associated with fluctuations of the corneoretinal potential known to influence the nystagmus velocities recorded using ENG technology (Lightfoot, 2004). Unlike ENG recording systems, the calibration of VNG systems relies solely on the camera-eye distance and repeated recalibration during a test session is not required. The VNG method requires the patient's eyes to remain open, and caloric nystagmus was recorded in total darkness, achieved by the use of light-excluding goggles within which the VNG cameras were housed. At the three centers using a water stimulus, the caloric test was contraindicated for some patients (e.g., those with a mastoid cavity or perforation; patients who had just undergone electrocochleography). Although a cool-only air caloric screening procedure was used in such patients, they were excluded from analysis in this study.

TABLE 3. Reasons for noncompletion of the bithermal caloric test and the number of cases, by test center

Reason	Center A	Center B (Air)	Center C	Center D	Total	Mono
Vomited/nausea/too unwell to continue	15	14	2	2	33	28
Perforation, grommet, or mastoidectomy (i.e., asymmetric anatomy)	11				11	1
Previous tympanoplasty	5		1		6	6
Anxiety after at least 1 irrigation	1	1	1		3	0
Patient declined water caloric after description	1			2	3	1
Post ECochG (employed cool air "Dundas Grant" system)	3				3	3
Could not get tympanometry seal so used cool air	2		1		3	3
Hypermobile or unusual tympanic membrane	4				4	4
Time constraints imposed by patient	1	1	1	1	4	4
Other/not stated	2		4		6	5
No. cases not completing bithermal test (sum of above)	45	15	10	6	76	55
Cases not completing a monothermal test	14	1	2	4	21	
Total no. cases	278	140	31	41	490	469

Mono is the no patients who underwent at least two caloric irrigations at one temperature and so would have been able to complete a monothermal test.
ECochG, electrocochleography.

Test Order

The caloric test procedure and contraindication policy of the BSA recommendation were strictly applied with the following exceptions: the first stimulus temperature was chosen either at random or alternated across patients (half received warm first, half received cool first), and the choice of first ear to irrigate was at the discretion of the investigator. In many cases, this was the "suspect" ear; in others it was arbitrary or selected according to local protocol. In all cases, an alternating ear sequence was used (RLRL or LRLR). It used to be believed that the four SPV results of the bithermal caloric test may be affected by a process of adaptation and that test order could influence the results. Indeed, one study (Furman & Jacob, 1993) suggested that a correction for this could be made. However, Lightfoot (2004) showed that such observations were explained by changes in the corneoretinal potential if ENG recalibration was not undertaken before individual irrigations and that no significant physiological adaptation was observed when VNG technology was used to record nystagmus. The test order used in the current study was chosen to exclude all doubt in this regard. A minimum of 6 min (typically 7 min) was allowed to elapse from the start of one irrigation to the start of the next. The equipment used for VNG recording was a VNG Ulmer by Synapsys, Marseille, France (centers A and C) or an ICS ChartR VNG by G N Otometrics, Schaumburg, IL (centers B and D).

RESULTS

Of the 490 patients tested, 414 (85%) completed the bithermal test. It is illuminating to identify the variety of reasons for noncompletion, and these are detailed in Table 3. Many of the reasons shown in Table 3 are associated with the use of water and in centers A, C, and D, a simple cool-only "Dundas-Grant" air system was used when appropriate. In other cases, the test procedure had to be abandoned because of patient nausea or anxiety but it is noteworthy that of the 76 patients who did not complete the bithermal test, 55 did undergo sufficient irrigations to allow a monothermal result to be obtained, so in only 4.3% were no useful results obtained. In Table 3, the "Mono" column gives the number of noncompleters who did provide valid caloric data at one temperature. For these patients, a monothermal caloric test would be ideal. It is also worthy of note that the 11 cases in center A that did not complete the bithermal test because of a perforation, grommet, or mastoid cavity, most underwent Dundas-Grant cool air calorics but because of potential anatomical asymmetry, only one of these was considered to have had a monothermal test.

Data Analysis

The results from a total of 414 patients in whom complete bithermal results were available underwent analysis. First, a cursory inspection of the data revealed that the warm stimulus was more effective than the cool. To investigate this we excluded all cases of significant CP or DP (>20%) and bilateral hypofunction (<8 °/sec) and then calculated the mean of all warm and all cool responses. These were 29.3 °/sec and 21.7 °/sec, respectively. A paired t-test showed this to be highly significant ($p < 0.001$). Although significantly different results at the two temperatures were found for the air stimulus, the warm and cool responses to an air stimulus were more similar (32.5 °/sec and 27.6 °/sec, $p < 0.001$) than those of a water stimulus (28.1 °/sec and 19.3 °/sec, $p < 0.001$). Careful recalibration of the water stimuli failed to identify a calibration problem as the reason for this observation.

Bithermal Caloric Test Results

Of these 414 patients, 118 (28.5%) had a CP exceeding 20% and 94 (22.7%) had a CP exceeding 25%. Five patients (1.2%) had bilateral hypofunction (of which two had total bilateral vestibular failure). For DP, 57 (14%) had a DP exceeding 20% and 32 (8%) had a DP exceeding 25%. Depending on a center's choice of diagnostic criteria, these are the cases that a monothermal test should attempt to identify. The distribution of CP values is illustrated in Figure 1.

Monothermal Caloric Asymmetry Criterion

To identify optimum values for the warm and cool MCA criteria using clinical decision analysis, we constructed an Excel spreadsheet to compute the false-positive and false-

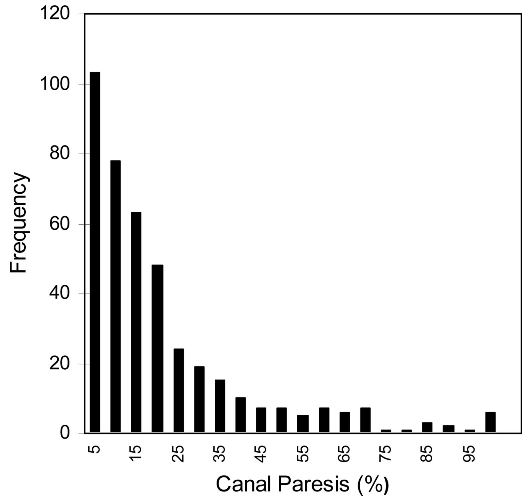

Fig. 1. Frequency distribution of canal paresis.

negative rates of the warm and cool monothermal tests using the three combined criteria discussed in the Introduction section. We started with our "best guess" values as follows: warm and cool MCA, 10%; warm and cool bilateral hypofunction, 8 °/sec; and spontaneous nystagmus, 3 °/sec. By calculating true-positive and false-positive rates for a range of warm and cool MCAs, receiver-operator characteristic (ROC) curves can be constructed (Turner & Nielsen, 1984). However, this does not provide the single figure we needed to identify the optimum criteria. The parameter d' does provide such a figure, but this gives equal weight to false-positive and false-negative errors. For our purpose, false-negative cases (missed cases of genuine CP or DP) are much more important than false positives, so we combined these two errors with a 10:1 weighting and then varied MCA from 4 to 25%. The MCA offering the lowest weighted error rate was chosen. When performing this analysis, we had to decide what values of bithermal CP and DP we should take as defining the "correct" answer. In recognition of the diversity of clinical practice, we performed the analysis twice for CP or DP = 20% and for CP or DP = 25%. After identifying optimum warm and cool MCA, we then turned our attention to the other two criteria (see below) and varied each in turn to select optimum values. Because we suspected that the three parameters were likely to interact and so influence the performance of the overall model, we performed this procedure iteratively until no change in performance was seen (three iterations for each parameter).

Absolute SPV Criterion—Evidence of Bilateral Hypofunction

The BSA Recommended Procedure offers a value of 8 °/sec as the minimum value for SPV. If all four SPVs for an individual are below 8 °/sec, a bilateral CP may be inferred. Although the bithermal test identified five patients in whom this criterion was satisfied, 6 (1.45%) and 11 (2.66%) patients satisfied the criterion if just their respective two warm or their two cool SPVs had to be below 8 °/sec. The greater number seen in cool tests is likely to be linked to the somewhat slower nystagmus generated by the cool stimulus. We therefore thought it likely that different warm and cool absolute SPV criteria would be appropriate when combined with the other two criteria to derive an optimal monothermal criterion.

Spontaneous Nystagmus Criterion—Evidence of DP

The spontaneous nystagmus seen in patients in the immediate aftermath of an acute vestibular incident can be viewed as a manifestation of their DP (Baloh, et al., 1977; Jacobson, et al., 1997; Shepard & Telian, 1996). As the process of central compensation occurs, both the nystagmus and the DP will diminish in most patients. Spontaneous nystagmus may therefore be taken as a surrogate measure of DP in most cases. Figure 2 illustrates the relationship we observed between DP and the mean value of nystagmus in the CTC when this was >3 °/sec. The high degree of correlation ($r = 0.86$, $p = 0.0001$) encouraged us to believe that it is reasonable to use the CTC nystagmus as one of our three criteria.

Overall Performance of the Monothermal Tests

After the multiple iterations of our analysis using different criteria for the three parameters, we were able to assess how well the two monothermal tests fared in predicting the outcome of the bithermal test. Our goal was to obtain sensitivity better than 95% (false-negative rate <5%) with at least 50% speci-

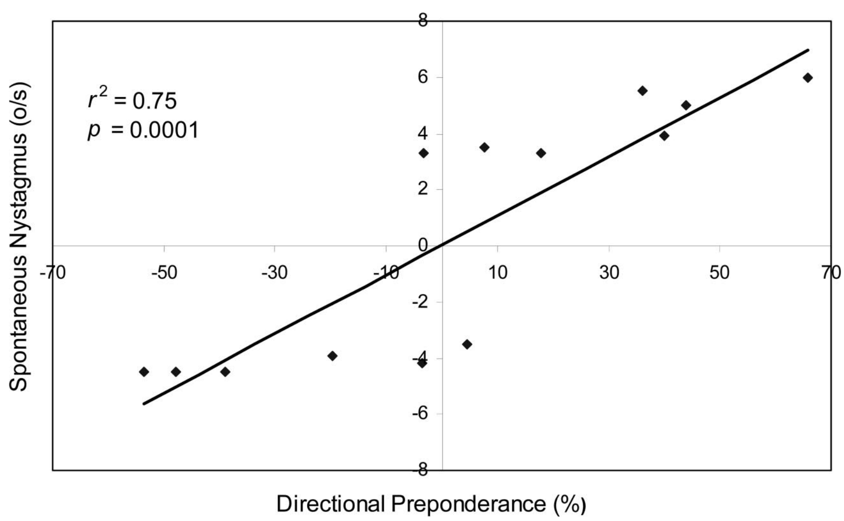

Fig. 2. The relationship between directional preponderance and nystagmus recorded in the caloric test condition.

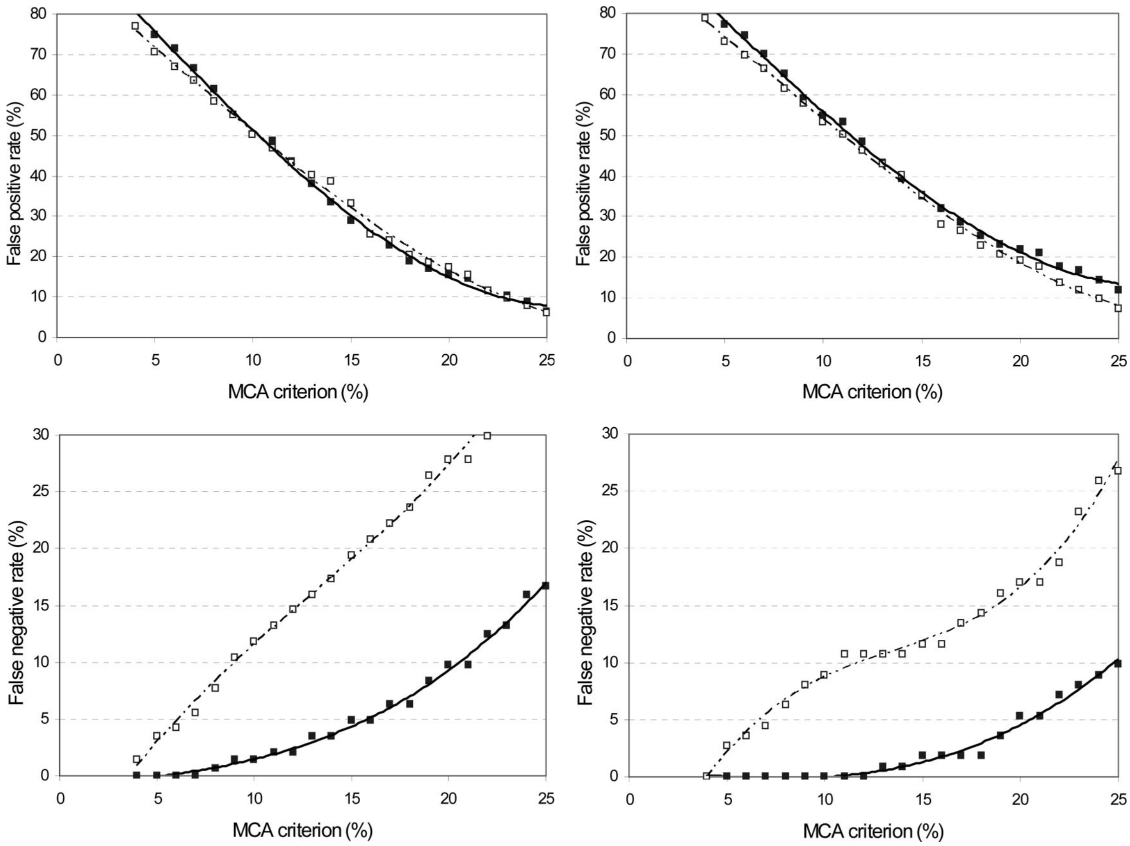

Fig. 3. Performance of the monothermal tests vs. the monothermal caloric asymmetry (MCA) criterion used. Left panels, 20% canal paresis (CP) and directional preponderance (DP) reference. Right panels, 25% CP and DP reference. Solid lines/filled symbols, warm test. Broken lines/open symbols, cool test.

ficity (false-positive rate <50%), thus sparing a useful proportion of patients with normal caloric test results from unnecessarily undergoing the full procedure. The optimum value for identifying a spontaneous nystagmus was 4 °/sec. For identifying bilateral hypofunction, the optimum warm value was 8 °/sec and for cool was 5 °/sec. These figures apply equally to centers adopting CP and DP criteria of 20 and 25%. The effect of the main parameter, MCA, is illustrated in Figure 3. The tradeoff between false-positive and false-negative rates for any given MCA criterion is clear. Also apparent is the marked difference in performance of the warm and cool monothermal tests. For the warm monothermal test, an MCA criterion of 15% is recommended. Taking 20% as the bithermal reference values for CP and DP, this criterion yields a sensitivity of just over 95% with only 29% false positives. Using a 25% bithermal reference for CP and DP yields a warm monothermal sensitivity of over 98% with 35% false positives. Recall that in this context, false positives are identified as true normals after receiving the two irrigations at the second temperature, at which time the bithermal results are available. Thus, the warm monothermal test offers a very attractive performance. For the cool test, the optimum MCA criterion was 7%. Here, performance is similar for both 20 and 25% bithermal CP and DP references. A sensitivity of around 95% is available, but at a high price: a false-positive rate of around 65%, offering the curtailment of testing to a relatively small proportion of patients.

We considered it possible that the superiority of the warm monothermal test might be associated with the faster nystagmus resulting from the warm stimuli. We scaled our warm and cool nystagmus results so that they generated the same mean value and using a 20% CP and DP reference, reanalyzed the warm and cool monothermal test performance, again identifying the optimum MCA criteria. This had negligible effect on the performance of the cool monothermal test but degraded the performance of the warm test somewhat (to retain sensitivity of 95%, the optimum MCA criterion became 12% and the false-positive rate was increased from 29 to 51%). Nevertheless, the warm monothermal test still outperformed the cool by a clear margin, suggesting that this difference is not solely explained by the relative efficacy of the warm and cool stimuli.

Water Versus Air as Stimuli

Not only do our warm stimuli generate faster nystagmus than the cool but also this warm/cool difference is more marked with water than air, despite careful calibration of the stimulus parameters. These characteristics have been described elsewhere (Zapala, et al., 2008). This raises the question whether our conclusions are equally valid for water and air media. To investigate this, the Liverpool (278 cases, water) and Windsor (140 cases, air) data were analyzed separately and their warm and cool false-positive and false-negative rates computed for a range of MCA criteria as before. This allowed examination of the relative test performance arising from the use of these stimuli. Analysis of variance revealed the expected temperature effect ($p < 0.0001$) in the false-negative performance. How-

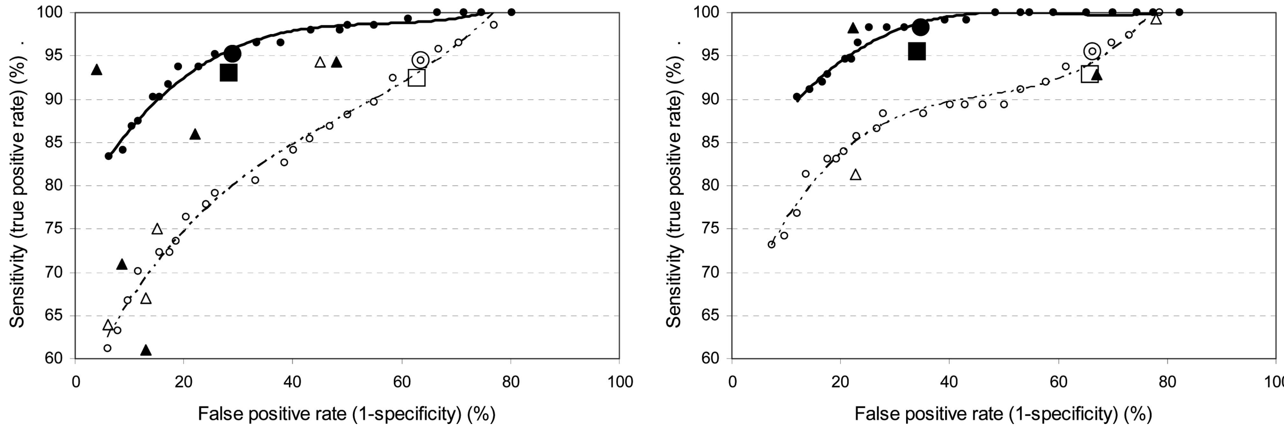

Fig. 4. Receiver-operator characteristics for the monothermal tests. Left panel, 20% canal paresis (CP) and directional preponderance (DP) reference. Right panel, 25% CP and DP reference. Solid lines/filled symbols, warm test. Broken lines/open symbols, cool test. Small circular symbols, data from this study. Large circular symbols, operating points suggested in this study. Square symbols, performance at the suggested operating points but without the spontaneous nystagmus and bilateral CP criteria. Triangular symbols, data from other studies (see Table 2).

ever, there was no significant ($p = 0.36$) effect of stimulus medium itself, but there was a temperature/medium interaction ($p = 0.0012$). Warm water provided the best performance; the warm/cool difference was less marked with the air stimulus. There were no noticeable effects of temperature or medium on the false-positive test performance. However, calculation of the optimum MCA criteria led to the conclusion that the stimulus medium was irrelevant: 15% was the optimum criterion for both water and air warm monothermal tests. Between 6 and 8% was the optimum criterion for both water and air cool monothermal tests (reduced sample sizes reduced precision so the range of optimum criteria was wider).

Comparison With Other Studies

Figure 4 gives the receiver operating characteristic of our three-element monothermal tests for 20 and 25% CP and DP reference values. The lines are ROC curves based on trend lines derived from this study's data. An ROC curve is a graphical representation of the power of a test. The curve of an ideal test should follow the top and left axes; the closer a test's curve is to the top-left corner the more powerful it is. Again, it is evident that the warm test offers superior performance to that of the cool. The large circular symbols denote the points corresponding to the MCA values we recommend. The square symbols show how the tests perform when using the MCA criteria alone, without the spontaneous nystagmus and bilateral criteria. The contribution of these two criteria to test performance in the overall population in our study is modest, but they are valuable for the small number of patients for whom they are relevant. In Table 2, we provided false-negative and false-positive rates that have been quoted in a number of earlier studies. The triangular symbols in Figure 4 represent these studies.

DISCUSSION

When designing a screening test a benchmark is needed against which the screening test's performance can be judged. As with previous studies, we have used the bithermal caloric test as this standard. However, it is worth remembering that the bithermal caloric test uses criteria for CP and DP that are derived from knowledge of the distributions of these variables in the normal population (usually based on 95% confidence intervals). This method defines the specificity of the test, but there is no guarantee that adequate sensitivity (to identify pathological patients) is obtained. The use of a 95% confidence interval for CP carries the expectation that we will falsely label 5% of nonpathologic cases (in terms of vestibular asymmetry) as pathologic and conversely, we will inevitably fail to identify genuinely pathologic cases having only a modest caloric asymmetry. When assessing the monothermal test, we feel it is important not to overlook the fact that the bithermal test does not offer perfect performance. Although the bithermal test is used as the standard against which the monothermal test is judged, because of its imperfection it would be inappropriate to demand substantially superior performance of its screening counterpart, the monothermal test.

Although it would be ideal to have a monothermal test that offered near-perfect sensitivity and specificity, pragmatically, a screening test that can detect almost all abnormal cases, while subjecting less than half of patients to the full bithermal test would be clinically acceptable. Although the warm monothermal test meets this challenge, our data suggest that the cool test does not. The reason for this probably lies in the relationship between CP and DP as discussed in the introduction and illustrated in Table 1. Our data support this: 83% of patients with a significant (20% criterion) DP have an opposing CP, and we believe that it is this interrelationship that is responsible for the superiority of the warm monothermal test. In this regard, this study is in agreement with most previous work. The only other study to recommend a warm MCA of 15% was that of Longridge and Leatherdale (1980) and it is encouraging that, using a bithermal CP reference of 25%, they recorded the same false-negative rate (1.8%) as observed in this study. A notable exception to the warm/cool superiority is the study by Enticott et al. (2003) who, using an air stimulus, found that the cool monothermal test offered better performance. The reason for this disparity is elusive. We initially suspected that it may be possible that the water and air stimuli give rise to better warm and cool monothermal test performance, respectively, or that the efficacy of the warm and cool stimuli that we observed (regardless of delivery medium) may explain this observation. The greater efficiency of the warm (excitatory) stimulus compared with the cool (inhibitory) stimulus has been reported before and is sometimes referred to as Ewald second law (Ford & Stockwell, 1978; Sills, et al., 1977). A recent study (Zapala, et al., 2008) provided a detailed and comprehensive analysis of the statistical properties of air and water caloric induced nystagmus and identified a greater response to warm than cool stimuli. Confusingly, however, the opposite finding has also been reported (Hood, 1989). Our own analysis of the monothermal tests suggests that the warm test offers superior performance for both media and, importantly, the same MCA criteria may be used with validity for both media. We are nevertheless intrigued by the differences we observed between this and previous studies and intend to pursue this by further investigation.

No previous study has used both VNG to record nystagmus and clinical decision analysis to guide the selection of optimum monothermal test criteria. VNG recording carries the promise of freedom from errors associated with the calibration drift inherent in electronystagmographic recording (Lightfoot, 2004) and clinical decision analysis provides clinically optimized criteria free from bias. We therefore hope that this study advances our understanding of this subject and will be used to inform future clinical practice.

Implications for Clinical Practice

To illustrate how the monothermal test is likely to work in a clinical setting, consider 100 patients. If our data are representative, typically 35 will have a significant CP, DP (20% criteria), or bilateral hypofunction, whereas 65 will give normal results as defined by the bithermal caloric test. Using our recommended criteria, a warm monothermal test will identify 95% of these 35 (33 cases), together with 29% of the remaining 65 (19 cases). The test therefore "fails" a total of 52 patients, who then go on to complete the bithermal test. Two abnormal cases are missed, but it is likely that one of these will have a CP <25%. Approximately half of 100 (48) cases do not have to undergo the cool tests.

The cool monothermal test is less helpful. To retain 95% sensitivity, an MCA criterion of 7% is appropriate and again, we would expect to identify 33 of the 35 abnormal cases. This time the number of false positives is 42 (65% of 65) and we therefore fail a total of 75 patients, sparing only 25 from the bithermal test. Although we have recommended values for the monothermal test criteria, the choice of MCA criterion is up to the user—Figure 3 may be used to select the MCA criterion giving the preferred performance tradeoff.

In Table 3, we analyzed the reasons for noncompletion of the bithermal test. The most common reason was patient nausea or vomiting. This tends to be a reaction to the cumulative effect of repeated caloric stimuli. Of these 76 patients, 55 did complete two tests at one temperature. One attraction of the monothermal test, therefore, is the expectation of obtaining clinically useful results in a greater proportion of patients.

Summary

We have demonstrated that the warm monothermal caloric test offers sufficiently satisfactory performance for us to recommend its use in the routine clinical investigation of the dizzy patient. Thus, the two warm tests should be performed first and the MCA calculated. The bithermal caloric test should be completed by continuing with the two cool tests if any of the following are not satisfied:

- the warm MCA must be below 15%
- both warm SPV results must be >8 °/sec
- there must be no evidence of a potential DP in terms of average spontaneous nystagmus faster than 4 °/sec.

We found that the above combination yields a sensitivity of 95% and a false-positive rate of 29% (71% specificity).

The cool monothermal caloric test offers performance that is inadequate for normal clinical use.

Postscript

During the manuscript reviewing process, a further 200 patients successfully completed the water bithermal caloric test at one of our centers (Liverpool). This allowed our predicted implications for clinical practice, above, to be tested using an independent sample. Again using 20% reference criteria for CP and DP, about one third (34%) of patients had an abnormal CP and/or DP. The use of our suggested warm monothermal criteria yielded a 94% sensitivity (63 of 67) and a false-positive rate of 27% (36 of 133). Of the four cases who constituted the 6% false negatives, three had a CP in the range 20 to 25%. Half (101 of 200) would have been spared undergoing the full caloric test had our warm monothermal criteria been used. The cool monothermal test had the poor performance we expected; whereas sensitivity was acceptable (94%), the false-positive rate was high (60%). We feel that analysis of this independent sample validates our conclusions.

ACKNOWLEDGMENTS

The authors thank Dr. Jas Sandhu and Joanne Enticott for their constructive comments on draft manuscripts.

Address for correspondence: Dr. Guy Lightfoot, Department of Clinical Engineering, Royal Liverpool University Hospital, Prescot Street, Liverpool L7 8XP, United Kingdom. E-mail: g.lightfoot@liverpool.ac.uk.

Received November 26, 2007; accepted August 8, 2008.

REFERENCES

Baloh, R. W., Sills, A. W., & Honrubia, V. (1977). Patients with peripheral and central vestibular lesions. *Ann Otol Rhinol Laryngol Suppl, 85*, 24–30.

Becker, G. D. (1979). The screening value of monothermal caloric tests. *Laryngoscope*, *89*, 311–314.

British Society of Audiology Recommended Procedure. (1999). Caloric test protocol. *Br J Audiol*, *33*, 179–84. Available at: http://www.thebsa.org.uk/pubs.htm. Accessed October 1, 2008.

Enticott, J. C., Dowell, R. C., & O'Leary, S. J. (2003). A comparison of the monothermal and bithermal caloric tests. *J Vestib Res*, *13*, 113–119.

Farid, S. A., El-Abd, S. M., & Abou-Elew, M. H. (2003). Monothermal caloric test its value in assessment of vestibular function. *Int Congress Ser*, *1240*, 1319–1324.

Ford, C. R., & Stockwell, C. W. (1978). Reliabilities of air and water caloric responses. *Arch Otolaryngol*, *104*, 380–382.

Furman, J. M. R., & Jacob, R. G. (1993). 'Jongkees' formula re-evaluated: order effects in the response to alternate binaural bithermal caloric stimulation using closed-loop irrigation. *Acta Otolaryngol*, *113*, 3–10.

Hood, J. D. (1989). Evidence of direct thermal action upon the vestibular receptors in the caloric test. *Acta Otolaryngol (Stockh)*, *107*, 161–165.

Jacobson, G. P., Calder, J. A., Shepard, et al. (1995). Reappraisal of the warm caloric screening test. *Ann Otol Rhinol Laryngol*, *104*, 942–945.

Jacobson, G. P., & Means, E. D. (1985). Efficacy of a monothermal warm water caloric screening test. *Ann Otol Rhinol Laryngol*, *94*, 377–381.

Jacobson, G. P., Newman, C. W., & Kartush, J. M. (1997). *Handbook of balance function testing*. San Diego: Singular.

Jongkees, L. B. W., Maas, J. P. M., & Philipszoon, A. J. (1962). Clinical nystagmography. *Practica Otolaryngol*, *24*, 65–93.

Keith, R. W., Pensak, M. L., & Katbamna, B. (1991). Prediction of bithermal caloric response from monothermal stimulation. *Otolaryngol Head Neck Surg*, *104*, 499–502.

Lightfoot, G. R. (2004). The origin of order effects in the results of the bi-thermal caloric test. *Int J Audiol*, *43*, 276–282.

Longridge, N. S., & Leatherdale, A. (1980). Caloric screening tests. *J Otolaryngol*, *9*, 478–481.

Norre, M. E. (1987). Screening methods for caloric testing. *Clin Otolaryngol*, *12*, 161–166.

Shepard, N. T., & Telian, S. A. (1996). *Practical management of the balance disorder patient* (1st Ed). San Diego: Singular. ISBN: 1879105845.

Sills, A. W., Baloh, R. W., & Honrubia, V. (1977). Caloric testing: results in normal subjects. *Ann Otol Rhinol Laryngol suppl*, *43*, 7–23.

Turner, R. G., & Nielsen, D. W. (1984). Application of clinical decision analysis to audiological tests. *Ear Hear*, *5*, 187–194.

Zapala, D. A., Olsholt, K. F., & Lundy, L. B. (2008). A comparison of water and air caloric responses and their ability to distinguish between patients with normal and impaired ears. *Ear Hear, 29,* 585–600.