**Sensitivity and Specificity of the Monothermal Caloric Screening Test**

**Introduction**

***What is a caloric test?***

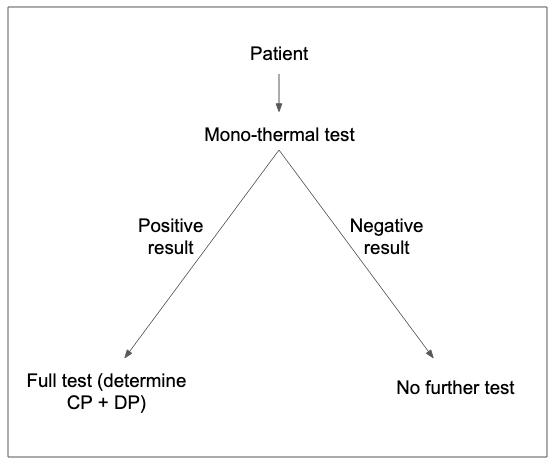
It is a test used to investigate the vestibular-ocular reflex (VOR), which is often impaired in balance disorders. This reflex automatically moves the eyes when the head is moved to enable a stable gaze. The caloric test involves introducing warm or cold air or water into the ear canal. This sets up a convection current in the semi-circular canal, which simulates a head turn. Heat will simulate a turn in one direction and cold in the other. In a patient with a normal VOR this will result in nystagmus and the extent to which the stimuli elicits nystagmus can be measured by recording peak slow-phase velocity (SPV). This procedure is done for each ear using both warm and cold stimuli and the resulting 4 measurements can be used to calculate Canal Paresis (CP), which is a measure of how different the two ears are, and Directional Preponderance (DP), which is a measure of difference in response between right and left simulated head turns. Additionally, the 4 measurements may indicate bilateral hypofunction, where the vestibular-ocular reflex is deficient in both sides and directions.

***What is a caloric test screen?***

The caloric test often has unpleasant side-effects for the patient, including nausea and vomiting. For this reason, limiting the number of people who need to undergo the full test will improve patient experience and save time. The full test involves measuring the effect of a single temperature (warm or cold) in both ears, then the alternative temperature in both ears. The recordings after a single temperature may give an indication of the full test results, so that if the single temperature results indicate a normal VOR, the patient does not need to undergo the measurements for the other temperature. The monothermal results may therefore be used as a screen for those patients which need to undergo the full test (Fig. 1).

***Objective***

This mini project uses a sample of patients who have undergone the full caloric test, to determine whether the results of the monothermal test can be used as a screen to reduce the numbers of patients undergoing the full test. In conducting a screen, there is a choice as to which criteria define a positive result in the screen. Choosing more lax criteria will increase sensitivity, at the expense of specificity and vice versa. This study will lay out the consequences of choosing different criteria.



**Fig. 1 Using the mono-thermal test as a screen.** If the screen result is positive, people will undergo the full test. If the screen result is negative, they will undergo no further test.

**Materials and Methods**

***Subjects***

800 patients undergoing the full caloric test were included in the study and each value for peak nystagmus slow phase velocity (SPV) was recorded. The order of testing (warm first or cold first) was also recorded. All patients completed all 4 measurements so there was no missing data. No additional information was provided about the subjects.

***Statistical software***

All statistics were performed in R. ROC analysis was performed using the ROCit package. Regression modelling was performed using the RMS model.

***Full test results in study population***

In order to compare the screen results with the full test results and determine specificity and sensitivity, the results of the full test were first determined. CP was calculated using the formula:

((WR+CR) - (WL+CL)) / (WR+WL+CR+CL)\*100%

DP was calculated using the formula:

((WR+CL) - (WL+CR)) / (WR+WL+CR+CL) \*100%

A CP or DP result >20% was considered a significant finding. Additionally, any case where SPV <8°/sec for all 4 measurements was considered significant (14 cases, 1.75%). These are both criteria used by Lightfoot et al. (2009). The prevalence of any significant finding upon the full test in this population of subjects was 35.25%.

***Monothermal screening test values***

Monothermal screen values were calculated using the warm values, for patients who had the warm test first (418), and cold values for patients who had the cold test first (382).

Warm screen values were calculated using the formula:

(WR-WL)/(WR+WL) \*100%

Cold screen values were calculated using the formula:

(CR-CL)/(CR+CL) \*100%

Any cases where the two SPV readings (WR and WL or CR and CL) were <8 were recorded as significant findings regardless of monothermal screen score and for this reason their screen score was set to NA, so they were excluded from the model.

***Modelling the relationship between positive test result, MS score and temperature***

Before doing the ROC analysis, a logistic regression model was created. This creates a ‘theory’ of how MS score and test temperature relate to a significant finding. The ROC analysis was then conducted on the output from the model. More detail about the model is included in the appendix.

The model used was as follows:

f <- lrm(gc\_binary ~ rcs(MS, 3) \* Order, data=d)

**Results and discussion**

***Sensitivity and specificity of mono-thermal tests***

The ability of the mono-thermal screen to discriminate between true positives (a patient with any significant finding) and true negatives is shown as a ROC chart in Figure 2. An AUC of 0.92 indicates the mono-thermal screen is a good indication of the true full test result.





**Figure 2. Mono-thermal test ROC analysis.** **(Top)** ROC curve showing an AUC of 0.93 (0.91-0.95 95% CI). **(B1)** The false positive rate for each possible cutoff between 0 and 30. **(B2)** The false negative rate for each possible cutoff between 0 and 30. Note the cutoffs here are probabilities generated by the model rather than MS values.

The table below and figure 2 show the specificity and sensitivity of the monothermal screening test in our sample, when findings are significant if either the screen value is above the specified cutoff or both STR values are <8. Table1 also shows how different cutoffs relate to different numbers of false negatives, false positives and numbers of patients having to undergo the full test, given a true positive prevalence of 33%.

For example, this shows that using a cutoff of 0.13 (model probability) will result in 3.3 false negatives (missed diagnoses) in every 100 patients seen (assuming a prevalence of 35.1% true positives). The consequences of missing a diagnosis could be serious for patients so it is desirable to use a cutoff that gives the test greater sensitivity. At the opposite extreme, using a cutoff of 0.035 (model probability) gives the test a sensitivity of 99%, however this means 87% of patients end up requiring the full test, which may mean there is no significant benefit in screening.

The best solution is subjective and depends on the seriousness and cost of missing a diagnosis vs the cost (in terms of both staff time and patient wellbeing) of performing the full test unnecessarily. It is suggested in the (Lightfoot et al., 2009)study that false negatives should be considered to be 10x more costly than false positives and for this reason if the test has 95% sensitivity then down to 50% specificity can be tolerated. In Table 1 the closest cutoff to this scenario is 0.077 (model probability). 95% sensitivity is achieved, with 58% specificity. Although the specificity is not very high, this still saves over a third of patients from undergoing the full caloric test.

**Table 1. Table showing implications of choosing different cutoffs for the screening test.** A lower cutoff, reduces numbers of false negatives but means large numbers of patients undergo the full screen. FN = false negatives, FP = false positives, NPV = negative predictive value, PPV = positive predictive value. Note that the cutoff here corresponds to a predicted probability from the model, which corresponds to different MS cutoffs for warm and cold tests (WarmMS, ColdMS).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| WarmMS | ColdMS | Cutoff | Sensitivity | Specificity | FN | FP | TN | TP | NPV | PPV | FT |
| 17.32 | 10.89 | 0.13 | 0.9 | 0.72 | 3.26 | 18.41 | 48.43 | 29.9 | 0.94 | 0.62 | 48.3 |
| 14.29 | 6.69 | 0.077 | 0.95 | 0.58 | 1.7 | 28.2 | 38.64 | 31.46 | 0.96 | 0.53 | 59.66 |
| 8.72 | 1.96 | 0.043 | 0.98 | 0.33 | 0.65 | 44.65 | 22.19 | 32.51 | 0.97 | 0.42 | 77.15 |
| 5.14 | 0.48 | 0.035 | 0.99 | 0.19 | 0.39 | 54.18 | 12.66 | 32.77 | 0.97 | 0.38 | 86.95 |

***Performance of the warm screen vs cold screen***

The two kinds of monothermal screen were analysed separately (418 warm, 382 cool). ROC analyses show better performance for the warm screen than the cold screen (Fig. 3). As the cutoff is moved from 0 to 0.5 (model probability), the warm screen displays a lower false negative rate and lower false positive rate. Using the warm screen with a cutoff of 12.29 gives 95% sensitivity and 67% specificity, which meets the parameters defined as acceptable by Lightfoot et al (2009, Table 3). This saves nearly half of patients from undergoing the full screen. Using the cold screen there is no cutoff which meets these criteria.

****

****

**Figure 3. ROC analyses for warm and cold monothermal screens. (Top)** The AUC for the warm screen is 0.95 (0.92 - 0.98 95% CI) and the AUC for the cold screen is 0.88 (0.84 - 0.92 95% CI). **(B1)** The false positive rate for each possible cutoff between 0 and 30 for cold and warm screens. **(B2)** The false negative rate for each possible cutoff between 0 and 0.5 (model probability) for cold and warm screens.

**Table. 2 The implications of different cutoffs for the cold monothermal screen**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ColdMS | Cutoff | Sensitivity | Specificity | FN | FP | TN | TP | NPV | PPV | FT |
| 8.63 | 0.098 | 0.9 | 0.48 | 3.29 | 35.06 | 31.78 | 29.87 | 0.91 | 0.46 | 64.94 |
| 7.64 | 0.087 | 0.95 | 0.43 | 1.79 | 37.8 | 29.04 | 31.37 | 0.94 | 0.45 | 69.17 |
| 5.48 | 0.066 | 0.98 | 0.33 | 0.6 | 44.93 | 21.91 | 32.56 | 0.97 | 0.42 | 77.49 |
| 3.64 | 0.053 | 0.99 | 0.21 | 0.3 | 52.6 | 14.24 | 32.86 | 0.98 | 0.38 | 85.46 |

**Table 3. The implications of different cutoffs for the warm monothermal screen**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| WarmMS | Cutoff | Sensitivity | Specificity | FN | FP | TN | TP | NPV | PPV | FT |
| 21.38 | 0.291 | 0.9 | 0.94 | 3.25 | 4.24 | 62.6 | 29.91 | 0.95 | 0.88 | 34.15 |
| 12.29 | 0.059 | 0.95 | 0.67 | 1.62 | 22.2 | 44.64 | 31.54 | 0.96 | 0.59 | 53.73 |
| 4.85 | 0.035 | 0.98 | 0.3 | 0.7 | 46.89 | 19.95 | 32.46 | 0.97 | 0.41 | 79.35 |
| 1.49 | 0.031 | 0.99 | 0.1 | 0.23 | 60.36 | 6.48 | 32.93 | 0.97 | 0.35 | 93.28 |

**Limitations**

It is important to highlight the limitations of this analysis. There is no additional information regarding the patients included in this sample, and it is assumed they are representative of patients seen by the clinic in general. The Lightfoot et al (2009) study also considers a third criteria, spontaneous nystagmus in the pre-test condition to be included. This study has no information on this, but inclusion of this information, may change the sensitivity and specificity of the screen. Patients who showed SPV readings < 8, suggestive of bilateral hypofunction were excluded from the analysis

**Conclusions**

This analysis indicates monothermal caloric test results can be used as a screen to save some patients from undergoing the full test. It indicates that performing the warm measurements first improves the sensitivity of the screen, a result which was also found in a previous study (Lightfoot et al., 2009). The implications of choosing different cutoffs are detailed, with the cutoff for the warm monothermal screen of 12.29 providing sensitivity and specificity at acceptable levels.

**References**

Lightfoot, G., Barker, F., Belcher, K., Kennedy, V., Nassar, G., Tweedy, F., 2009. The Derivation of Optimum Criteria for Use in the Monothermal Caloric Screening Test. Ear Hear. 30, 54–62. https://doi.org/10.1097/AUD.0b013e31818f006c