Informatics in Primary Care 2010;18:171–5

Refereed paper

Using keywords to predict the need for an audiogram: an analysis of referral letters using logistic regression

Venkat M Reddy MRCS DOHNS
Specialty Registrar in Ear, Nose and Throat (ENT)

Tarig Abdelrahman BM
Senior House Officer (CT1) in ENT

Andrew Lau MRCS
Senior House Officer (CT2) in ENT

Kel Anyanwu FRCS
Associate Specialist in ENT

David JD Whinney FRCS (ORL-HNS)
Consultant ENT Surgeon

ENT Department, Royal Cornwall Hospital, Truro, Cornwall, UK

ABSTRACT

Background Otolaryngology clinics are often booked without considering the distribution of work for doctors and audiologists. This causes inefficiencies of time and human resources. This may be due to clinics being booked before referrals have been triaged to identify whether a hearing test, known as a pure tone audiogram (PTA), is indicated. A model that can predict the need for PTA without clinician-led triage could be useful to address these booking issues.

Objective To establish if it is possible to predict whether a referred patient requires a PTA based on occurrences of words in the referral letter.

Method Binary logistic regression analysis of 500 letters of referral for otolaryngology outpatients. The derived model was then tested on 50 referral letters. All the referral letters were reviewed by clinicians and classified according to whether or not a PTA would be required.

Results The regression model correctly predicted requirement for a PTA in 42 of 50 referral letters (84%), with a sensitivity of 91% and specificity of 82%.

Conclusion The model is able to predict requirement for a PTA from referral letters with a high degree of accuracy. This method may have a role in assisting administrative/clerical staff or non-specialist clinicians to book appropriate ear, nose and throat (ENT) clinic appointments, with or without a PTA. As a result, workload would be distributed more evenly, through the clinic for both otolaryngologists and audiologists, increasing efficiency.

Keywords: audiometry, informatics, otolaryngology, outpatients, referral and consultation

Introduction

General practitioners make referrals for otolaryngology consultations through correspondence, which may be in paper-based or electronic form, through email or Choose and Book. Whereas traditionally a paper-based referral letter was triaged by the clinician to determine whether a hearing test, otherwise known

as a pure tone audiogram (PTA), was required, the vast majority of referrals are now received through Choose and Book. Clinic appointments are booked before the referrals have been reviewed by an ENT doctor or specialist nurse. The process of checking referrals allows for the identification of patients for whom a PTA is indicated. It is important to ensure that these patients are booked into clinics that have audiology services available. The booking of clinic slots before the referral has been reviewed may lead to an uneven distribution of cases on clinic lists requiring PTAs. This leads to inefficient use of audiologist and clinician time in clinic, with consequences for the timely running of clinics. We wanted to explore whether we could predict which referrals would require a PTA based on the occurrence of discriminating words in referral letters. This could form the basis for producing a software-based tool capable of analysing the text in a referral letter to inform administrative staff (in both primary and secondary care) of the likelihood of a PTA being required. This would assist in booking clinics with a more even distribution of work for clinicians and audiologists, thus improving clinic efficiency.

Objective

To establish whether referrals requiring a PTA at the initial clinic appointment can be predicted by a model based on word frequencies in the referral letter.

Materials and methods

The study involved 500 consecutive letters of referral for outpatient consultations received by the ENT department of the Royal Cornwall Hospital through the electronic referral system known as Choose and Book during October and November 2009. Only the main body of text was included. All patient identifiable information was discarded in order to maintain confidentiality. The letters were reviewed by the authors and classified according to audiogram requirement in binary terms, with 0 implying that PTA was not indicated and 1 implying that PTA was indicated. The main text of the letters was analysed using TextSTAT 2.8 software (Matthias Hüning, Berlin, Germany) to identify the words that occurred most frequently. The authors identified words thought to be most relevant in distinguishing whether an audiogram was indicated. A Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, Washington, USA) was designed to count the occurrence of the relevant words in each referral letter. Logistic regression analysis using a forward stepwise conditional method was performed using SPSS 11.0 (SPSS Inc., Chicago, Illinois, USA). PTA requirement was considered as the dependent variable, and words in the referral letter were considered as explanatory variables. The resulting logistic regression model predicted whether the referred patients required a PTA based on frequencies of relevant words in the referral letters.

A further 50 consecutive referral letters received by the department were also analysed by an ENT surgeon and allocated to either group. Data was collected on a Microsoft Excel (Microsoft Corp., USA) spreadsheet incorporating formulae to count relevant word frequencies and formulae to apply the regression model. The resulting information was used to establish the sensitivity and specificity of the model.

The study protocol was reviewed by the local ethics committee and the institutional research review body and was deemed as not requiring ethical approval.

Results

Five-hundred consecutive referral letters were analysed by an ENT surgeon, of which 255 referrals required an audiogram (group A) and 245 did not (group B). Text analysis identified 4461 different words throughout the main body of text of all the referral letters. The first two authors (VR and TA) identified 150 words as potentially relevant for discriminating between groups A and B. Binary logistic regression analysis using a forward stepwise conditional method was performed. The resulting model correctly predicted the indication for audiometry in 469 of 500 referral letters (94%, Nagelkerke R square=0.88, chi square=542.44, $P<0.001$). The model consisted of 23 words and a constant. The words and the corresponding coefficients are shown in Table 1. Details on the accuracy of the model are shown in Table 2. The format of the regression model is shown in the following equation:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots\ldots\ldots \beta_{24} x_{24}$$

The components of the regression model are as follows: B_0 is the constant –2.2422876533113); B_1 represents the coefficient that applies to a word in the model (see Table 1) and x_1 represents the frequency of that word in the referral letter. The derived value, z, is applied in the following equation:

$$y=e^z$$

If $y=\leq 1$, the model predicts that audiometry is not indicated. If $y=\geq 1$, the model predicts that audiometry is indicated.

Table 1 Components of the logistic regression model with words in descending order of influence

Components of the model	Beta coefficient (standard error)	Lower 95% CI for odds ratio	Odds ratio	Upper 95% CI for odds ratio
CONSTANT	–2.242 (0.355)			
Membrane	12.753 (3.046)	882.81	345 425.93	135 157 729.73
Dizziness	7.794 (1.727)	82.18	2426.72	71 655.31
Auditory	6.230 (1.858)	13.31	507.63	19 367.04
Hearing	5.154 (0.850)	32.73	173.14	916.04
Sore	4.433 (1.510)	4.36	84.16	1623.52
Deaf	3.928 (1.101)	5.87	50.80	439.39
Behind	3.681 (1.267)	3.31	39.70	475.80
Vertigo	3.313 (1.105)	3.15	27.46	239.45
Hear	3.063 (1.229)	1.92	21.40	238.16
Perforation	2.346 (1.194)	1.01	10.44	108.45
Ear	2.169 (0.382)	4.14	8.75	18.48
Tinnitus	1.829 (0.609)	1.89	6.23	20.53
Months	0.974 (0.484)	1.03	2.65	6.84
Nasal	–1.359 (0.544)	0.09	0.26	0.75
Wax	–2.440 (0.693)	0.02	0.09	0.34
Clear	–2.640 (1.024)	0.01	0.07	0.53
Externa	–3.057 (0.943)	0.01	0.05	0.30
Throat	–3.452 (1.339)	<0.01	0.03	0.44
Long	–4.026 (1.339)	<0.01	0.02	0.25
Dull	–5.299 (2.304)	<0.01	<0.01	0.46
Tonsillectomy	–5.658 (1.659)	<0.01	<0.01	0.09
Aid	–7.045 (1.364)	<0.01	<0.01	0.01
Morning	–8.773 (1.815)	<0.01	<0.01	0.01

R^2 values for model: Nagelkerke R^2=0.883, Cox and Snell R^2=0.662

Table 2 Classification table showing predicted and observed results of 500 referral letters used in the binary logistic regression analysis

		Predicted (by model)		
		Audiogram	No audiogram	Totals
Observed	Audiogram	234	21	255
	No audiogram	10	235	245
	Totals	244	256	500

Table 3 Classification table showing predicted and observed results in 50 referral letters not used in the binary logistic regression analysis

		Predicted (by model)		
		Audiogram	No audiogram	Totals
Observed	Audiogram	24	4	28
	No audiogram	4	18	22
	Totals	28	22	50

Fifty referral letters were subjected to the regression model which correctly identified the requirement for audiometry from 42 of the 50 letters (84%). The model demonstrated a sensitivity of 91% and a specificity of 82% (see Table 3).

Discussion

The regression model that was derived correctly predicted the requirement for audiological assessment in 84% of referral letters. The model demonstrated a high sensitivity (91%) and specificity (82%). This is not as accurate as having a clinician triaging referrals, but the high degree of accuracy of the model is surprising considering that it is based only on word frequencies. It is possible that the application of more advanced text processing algorithms may improve the accuracy of the model.[1]

The use of text analysis in the management of healthcare information has the potential to improve healthcare communication processes. Esquivel *et al* developed a statistical model that was able to predict the acceptance of 74.7% of referrals made by primary care providers to specialists based on the total word count of the referral letter.[2] Such a model could be incorporated into software, facilitating communication between clinicians in order to estimate the likelihood that the referral will be declined and prompting the referrer to provide more information.[2]

The model that we have derived may be useful to non-clinical administrative staff, with little or no knowledge of otolaryngology, with a role in the booking of outpatient appointments. The booking of clinic appointments prior to clinician review of referrals is unavoidable as referral-to-treatment times shorten, and as uptake of Choose and Book electronic referral increases.[3] Use of the Choose and Book system can be time consuming for a busy general practitioner so the task is often delegated to administrative staff in the practice.[4] Availability of a software-based tool employing the model that we have derived will allow administrative staff to identify most of the referrals that require a PTA. They can proceed to distribute the appointments evenly throughout a clinic in order to avoid clustering of patients requiring audiograms. This may help to make better use of audiologist and clinician time, allowing for the timely running of the clinic. This process is not intended to replace formal review of referrals by a clinician. Further investigation is required to establish whether such a model is acceptable to non-clinical administrative staff, and to see if it would bring about objective improvements in the running of clinics.

Conclusion

The logistic regression model is able to predict from referral letters, with a high degree of accuracy, the requirement for a PTA. This function may have a role in assisting administrative/clerical staff or non-specialist clinicians to book appropriate ENT clinic appointments, with or without a PTA. As a result workload would be more evenly distributed through the clinic for otolaryngologists and audiologists, increasing efficiency.

REFERENCES

1 Cohen KB and Hunter L. Getting started in text mining. *PLoS Computational Biology* 2008;4:e20.
2 Esquivel A, Dunn K, McLane S, Te'eni D, Zhang J and Turley JP. When your words count: a discriminative model to predict approval of referrals. *Informatics in Primary Care* 2009;17:201–7.
3 Walford S. Choose and Book. *Clinical Medicine* 2006; 6:473–6.
4 Rabiei R, Bath PA, Hutchinson A and Burke D. The National Programme for IT in England: clinicians' views on the impact of the Choose and Book service. *Health Informatics Journal* 2009;15:167–78.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Venkat M Reddy
ENT Department
Royal Cornwall Hospital
Truro
Cornwall
TR1 3LJ
UK
Tel: +44 (0)1872 253405
Email: vmreddy9@gmail.com

Accepted September 2010