Predicting the effects of physician-directed promotion on prescription yield and sales uptake using neural networks

Received (in revised form): 23rd April, 2004

Chee Wooi Lim

earned his degree from the Graduate School of Pharmaceutical Sciences, Kyoto University. He has previously worked with the development and application of computational methods to problems in drug delivery research. Much of his work was centred on pharmacokinetic analysis of drug disposition in skin and prediction of skin permeability. Currently, he is working at the International Marketing Department in Daiichi Pharmaceutical Co., Ltd. His role combines market research with special interest in promotional effectiveness.

Toru Kirikoshi

is a manager at the International Marketing Department of Daiichi Pharmaceutical Co., Ltd. He has 12 years' worldwide marketing experience in the field of antimicrobial agents.

Abstract In this paper, the potential of neural networks to predict the effects of promotion on prescription yield and sales uptake is explored. The motivation for this work derives from the notion that marketing strategies guided by predictions of the effects of promotion help in identifying marketing expenditure that needs augmentation, thereby generating a healthier return on investment. Five promotional spending descriptors for an antibiotic product are selected as descriptive data (network inputs) to neural networks with prescription yield and sales uptake volumes as the outcomes (network outputs). A back-propagation neural network with five hidden elements (commonly called neurons) compares favourably with the benchmark obtained by multiple linear regressions in terms of r^2 and mean absolute percentage error (MAPE). Further, a 'leave-one-out' cross-validation revealed that the optimal neural network is able to correctly predict the effects of promotion more than 95 per cent of the time ($r^2 \geq 0.825$, $\text{MAPE} \leq 4.3$ per cent), suggesting the general utility of this approach as a potential tool for predicting the effects of physician-directed promotion.

INTRODUCTION

With spending on promotion the fastest growing component of the marketing budget for prescription drugs, concerns about quality and promotional effectiveness have grown over the past few years in the pharmaceutical industry.¹ Despite this fact, little detailed material is published about the methodological aspects of predicting the effects of

promotion. Prediction would clearly be valuable, but determination of its consequences seems difficult or impossible. To fully leverage a marketing engine, a manager must employ field-proven tools that could increase understanding of the relevance and limitations of his/her promotional spending. Specifically, a reliable prediction of the outcomes associated

Chee Wooi Lim
Daiichi Pharmaceutical Co.
Ltd, International Marketing
Department,
14-10, Nihonbashi
3-Chome, Chuo-ku, Tokyo
103-8234, Japan.
Tel: +81 3 3273 9179;
Fax: +81 3 3276 0694;
e-mail: rimuchoo@
daiichipharm.co.jp

with a wide range of spending levels is a prerequisite in helping to reduce unnecessary expenditure, as well as in identifying promotional expenditure that needs augmentation.

Many managers plan their promotional budget allocation based on experts' opinions or panel consensus.² The risk is that these experience-based judgments are too static to accommodate in time the changing dynamics of the marketplace that often require decisions to be quickly revised. The conventional quantitative approach, based on multiple linear regression analysis, while substantively appealing, is sometimes problematic in practice due to the difficulty of accurately assessing the interactive effects between variables. In common with many multiple linear regressions, a number of descriptors are found to be important variables together, and complex relationships must be accounted for by *a priori* inclusion of interaction terms.³ A more managerially useful approach is therefore to embrace a less stressful tool that could predict observed variation in the outcome variable without knowledge about the nature of the relationship between the explanatory factors and the outcomes, ie data go in and a prediction comes out. Neural networks are clearly well suited to meet this need.

In this study, some basic notions of neural network modelling have been extended to reveal the effects of promotion on prescription yield and sales uptake volumes. The objectives of this work are, first, to produce an understanding of the factors which affect prescription yield and sales uptake by sorting through the relationship in the data using neural networks. Secondly, the optimal neural network model is then called upon to provide a prediction of unknown prescription yield and sales uptake given a forecast of the key

factors. The paper begins by feeding the history of past prescription yield and sales uptake volumes together with several main promotional spending descriptors into neural networks. The major goal of this study is, however, to formulate a neural network model that makes an accurate prediction of the effects of promotion on prescription yield and sales uptake, and therefore subsequent analysis includes employing the optimal neural network to predict unknown prescription yield and sales uptake volumes with the aim of confirming the validity of this approach. To provide a more complete evaluation, the predictive ability of neural networks is compared with the benchmark obtained by multiple linear regressions.

MATERIALS AND METHODS

Data collection

Data used in this study are derived from a market-leading product (Brand A) in the US antibiotic market. These cover a 21-month period with five promotional spending descriptors as input layer, and prescription yield and sales uptake volumes as output layer. The data are normalised to protect the commercial competitive advantage of this product in the marketplace. A descriptive statistics summary for Brand A is presented in Table 1. In Table 1, 'contacts' (CNT) measures the number of visits (details) made by pharmaceutical representatives to physicians and 'cost of contacts' (COC) includes the costs associated with detailing of representatives that are directed to physicians. 'Minutes' (MIN) is the projected sum of time spent with physicians whereas 'advertising spending' (ADS) captures the expenditure of advertising in medical journals. 'Sample' (SMP) shows the projected volume of Brand A provided as samples to

Table 1: Descriptive statistics summary for Brand A

Descriptor	Designation	Mean	Minimum	Maximum	Standard deviation	<i>n</i> (months)
Contacts	CNT	68	54	79	8	21
Cost of contacts	COC	6,473	4,774	8,160	847	
Minutes	MIN	334	211	483	63	
Advertising spending	ADS	168	0	558	186	
Sample	SMP	1,256	980	1,635	190	
Prescription	P	959	783	1,195	108	
Sales	S	59,752	43,966	76,020	10,176	

physicians. The effects of promotion are reflected in the measurement of ‘prescription yield’ (P) that represents the count of Brand A prescriptions written by physicians and ‘sales uptake volumes’ (S) that captures purchases of Brand A by pharmacies.

Network configuration

A neural network is a set of simple elements (commonly called neurons), each linked to some of the others that transmit information to each other through these links (connections). Neural networks used in this study consisted of three structural layers (Figure 1): an input layer with five promotion descriptor variables, a hidden layer with varying number of neurons, and an output layer with two outcome variables. Before training was initiated, the input values were scaled to minimise the chance of saturation due to large input values. An error back-propagation algorithm with weight updates occurring after each epoch was used for network training, where the learning rate was set at 0.1. Learning goes as follows: each historical record of promotional spending is presented as input to the neural networks; these values are then propagated through the connections towards the outputs. The prediction obtained at the network’s outputs is sometimes erroneous (different from the actual output value presented). The error value is then computed and back-propagated by moving upwards in

the network and modifying the weights proportionally to each one’s contribution to the total error value. This mechanism is repeated for each historical record in the training set. One of the primary goals in training neural networks is to ensure that the networks will perform well on data on which they have not been trained; this is called ‘generalisation’. A 5 per cent level of Gaussian noise was therefore added to the training data to help prevent ‘overfitting’ and improve ‘generalisation’.⁴ The goodness-of-fit was evaluated via mean absolute percentage error (MAPE), defined as the arithmetic average of the percentage error the neural networks estimated:

$$\text{MAPE} = \frac{\sum_{i=1}^n |\%e_i|}{n}$$

where $|\%e_i|$ denotes absolute value of the percentage error in month i , and n represents number of months.

MAPE is used because of its popularity in the forecasting literature and because it is not prone to changes in the magnitude of the time series to be forecast.⁵ Since prescription yield and sales uptake tend to fluctuate over time, a measure of prediction error such as the root mean square error may be misleading when two periods are compared. Further, the rationale was supported by the fact that MAPE could be used to monitor forecasting model accuracy.⁶

Figure 1 Schematic representation of a three-layered neural network

The predictive ability of neural networks was evaluated by a 'leave-one-out' cross-validation procedure. This method systematically removes one data point at a time from the training set. A network model is then constructed on the basis of this reduced data set and is subsequently used to predict the removed data. This procedure is repeated for all data so that a complete set of predicted values is obtained.

RESULTS

In Figure 2, although there are monthly variations, it is notable that the overall patterns between sales uptake and prescription yield of Brand A are somewhat similar. For instance, in general, the peak sales and prescription months are during the 1st–3rd months and 10th–15th months. After the 1st–3rd months, the sales and prescriptions decline but there are some perturbations in the data with sales volumes rising slightly in the 6th and 18th months, then declining to their lowest point during

7th and 19th months. The reason is that sales data may not immediately reflect the actual physicians' prescribing pattern captured by prescription data. Despite this lagged response, the inclusion of sales data in the output layer was a conscious decision as the initial aim sought to make for a challenging prediction problem.

In Table 2, when multiple linear regression analyses of prescription yield and sales uptake volumes were performed against the promotional descriptors, the following equations were obtained, respectively:

$$P = 671 - 3.63CNT + 0.02COC + 0.70MIN - 0.31ADS + 0.18SMP$$

$$r^2 = 0.210, \text{ MAPE} = 8.4\%$$

$$S = 63,182 + 39.9CNT - 0.9COC + 44.1MIN - 13.8ADS - 10.0SMP$$

$$r^2 = 0.078, \text{ MAPE} = 14.3\%$$

Figure 2 Plot of sales and prescription for Brand A

Table 2: Results of multiple linear regression analyses of prescription yield and sales uptake volumes against five promotional spending descriptors

P	Coefficient	95% Confidence interval		t-test (P-value)
		Lower limit	Upper limit	
Intercept	671	169.0	1,173	0.012
CNT	-3.63	-29.9	22.6	0.772
COC	0.02	-0.251	0.293	0.873
MIN	0.70	-1.002	2.398	0.395
ADS	-0.31	-0.824	0.202	0.216
SMP	0.18	-0.224	0.575	0.365

S	Coefficient	95% Confidence interval		t-test (P-value)
		Lower limit	Upper limit	
Intercept	63,182	12,006	114,358	0.019
CNT	39.9	-2,637	2,717	0.975
COC	-0.9	-28.7	26.8	0.944
MIN	44.1	-129.2	217.4	0.596
ADS	-13.8	-66.1	38.5	0.582
SMP	-10.0	-50.8	30.7	0.608

where CNT, COC, MIN, ADS and SMP represent the contacts, cost of contacts, minutes, advertising spending and samples, respectively. The same data were subjected to neural network analyses.

Selecting a model size that maximises generalisation is an important topic. There are several theories for determining the optimal network size, such as the Network Information Criterion (NIC) which is a generalisation

Figure 3 MAPE and model complexity for a three-layered neural network with a 5- x -2 configuration ($x = 1, 2, \dots, 10$)

of the Akaike Information Criterion (AIC)⁷ widely used in statistical inference, the generalised final prediction error (GPE)⁸ and the Vapnik-Chervonenkis dimension⁹ which is a measure of the expressive power of a network. Nonetheless, the authors experimented with several different numbers of hidden layer neurons and found that the results were not very sensitive to the number of hidden layer neurons following these theories. Alternatively, they attempted to use part of the data to check the model in some way, such as cross-validation.¹⁰ Cross-validation is often used to select an optimal architecture from among a set of available network architectures. In a 1996 study cross-validation was found to be suitable for choosing the optimal network architecture, at least on the data sets tested.¹¹ In Figure 3, the mean absolute percentage error (MAPE) was calculated for neural network models

with a configuration of 5- x -2 ($x = 1, 2, \dots, 10$) in order to find the optimal one. The MAPE for sales was lowest (2.4 per cent) for a neural network model with a 5-5-2 configuration. On the other hand, the MAPE for prescription was lowest (1.7 per cent) for neural network model with a 5-7-2 configuration; however, the value was almost the same as that of a 5-5-2 model (2.1 per cent). In order to keep the model complexity at a minimum while preserving the model fidelity, the 5-5-2 neural network was adopted as optimal. Model complexity is expressed as the numbers of the model's free parameters, given by the number of connections between neurons.¹²

A critical test for the success of the neural network approach is to compare its performance for prescription yield and sales uptake prediction with the methods which are widely used in pharmaceutical marketing today. The multiple linear regression method was chosen because

Figure 4 Linear fits of predicted and actual prescription yield obtained from analyses using multiple linear regressions and 5–5–2 neural network

many managers are familiar with it. The plots in Figure 4 and Figure 5 show a linear fit for the predicted and the actual values of prescription yield and sales uptake volumes, based on both multiple linear regression and neural network analyses. A raw measure of performance on the predictive power by both models

can be measured via a common statistical measure, r^2 , where an r^2 of 1 indicates perfect prediction. The multiple linear regression models had an r^2 of 0.210 for prescription and 0.078 for sales which are much lower than those for the neural network models ($r_p^2 = 0.905$; $r_s^2 = 0.964$). Further, the neural network models also

Figure 5 Linear fits of predicted and actual sales uptake volumes obtained from analyses using multiple linear regressions and 5–5–2 neural network

outperformed multiple linear regression models from the viewpoint of MAPE (neural network: $\text{MAPE}_p = 2.1$ per cent, $\text{MAPE}_s = 2.4$ per cent; multiple linear regression: $\text{MAPE}_p = 8.4$ per cent, $\text{MAPE}_s = 14.3$ per cent) showing that the latter could not achieve a comparable predictive power to the former.

The optimal neural network model was used to predict values for unknown prescription yield and sales uptake volumes using a leave-one-out cross-validation procedure (Figure 6). The MAPE for the leave-one-out prediction analysis was around 4 per cent ($\text{MAPE}_p = 3.5$ per cent, $r_p^2 = 0.825$;

Figure 6 Linear fits of predicted and actual prescription yield, and sales uptake volumes in the leave-one-out cross-validated prediction

$MAPE_s = 4.3$ per cent, $r_s^2 = 0.909$), showing that the neural network model is able to correctly predict the effect of promotion more than 95 per cent of the time. This implies that there is a less than 5 per cent chance that these forecasting results may be due to chance alone.

DISCUSSION

Pharmaceutical companies employ a number of promotional strategies for prescription drugs designed to target physicians and consumers respectively. Because physicians have long been the key decision makers when it comes to choosing a prescription drug,

pharmaceutical companies traditionally have concentrated most of their marketing efforts on physicians, and still do so.¹³ For many companies, physician-directed promotion continues to command a large share of marketing budgets with transnational big firms investing billions of dollars a year. Among the top 14 pharmaceutical companies, the average return on investment of physician-directed promotion in the US market declined from \$22.2 to \$17.0 between 1998 and 2001.¹⁴ Since the relationship between investment and return is no more than linear in pharmaceutical promotion, profit margins will not increase by merely spending more. To improve the situation, pharmaceutical companies must establish an approach that could pinpoint the outcomes associated with a wide range of promotional spending levels before field test. For example, what would the prescription yield be if advertising spending were increased by 10 per cent? Similarly, one might want to learn about the mix of contacts and sample volumes that would be needed to achieve rapid sales uptake. Predicting what would happen under a hypothetical plan is particularly important. A small improvement in prediction can make a significant difference in operational efficiency or in bottom-line profits. For example, research has shown that winning just one more prescription per week from each physician could yield an annual gain of US\$52m in sales.¹⁵

In this study, the authors demonstrated that neural networks are able to predict the impact of physician-directed promotion correctly without an appreciably smaller r^2 value (Figures 4 and 5) than multiple regression analysis. Perhaps more importantly, the novelty of this work has another implication. For instance, if the impact of promotion can be easily measured, then the efficacy of

interventions to reduce the harmful consequences of skyrocketing drug expenditure could probably be tested.

Conventionally, most managers prefer multiple linear regression schemas as a shortcut for examining the effect of promotion because they are faster and simpler. Because choosing among a number of important explanatory variables associated with overall evaluations is presumed to be a thorny task, the neural network approach is considered less stressful for managers in that the complex interrelationship among explanatory variables can be isolated. Although introducing the interrelation terms between the descriptors in multiple linear regression analysis could circumvent this problem, it should be noted that multicollinearity of the descriptors decreases the robustness of a multiple linear regression model and makes the contribution of the descriptors unclear. On the other hand, neural networks can successfully take into account information involving the combined effect of several descriptors.

In common with all modelling studies, concerns may be expressed as to the selection criteria of promotional descriptors. For the five promotional descriptors chosen in this study, all that can be deduced is that they appear to be an understandable choice as the vast majority (more than 80 per cent) of spending on physician-directed promotion was concentrated on samples, detailing and advertising in medical journals.¹⁶ The choice of descriptors also builds on the fact that cumulative spending on both medical journals and detailing increases sales.¹⁷

Since this study is examined in a very homogeneous context, the wider applicability of its findings could probably be questioned. In particular, the study is comprised of only Brand A within the antibiotic class. It may

produce a spuriously optimistic outcome when tested but fail in practice if a product other than Brand A is substituted for testing. Simply Brand A was selected for this study, in part, because of its familiarity. Although it is not believed that the results would apply directly to the universe of products, two reasons suggest that this approach is valuable as a useful tool for predicting the effects of promotion on prescription yield and sales uptake for a considerable array of therapeutic classes. First, with spending on physician-directed promotion growing substantially over the past ten years, rising 60 per cent from 1996 to 2000,¹⁸ it is obvious that no effective strategy is being adopted, because measuring the impact of promotion seemed virtually impossible — let alone making the necessary changes. Secondly, considering that the neural network approach, as used in this study, will operate without any knowledge about the nature of the relationship between the outcome and the explanatory variables, prospective examination to adapt to changing circumstances by most managers appears to be warranted. On the other hand, it would be particularly relevant for brand-specific analyses, which the authors hope to investigate in the near future.

For one thing this study is cross-sectional in nature. As such the information is captured at only one point in time. Future research will investigate the possibility to control for extraneous sources of error that may have been introduced by the idiosyncrasies found across different time periods via a longitudinal study.

CONCLUSION

This paper has considered an important application of neural networks to predict the effects of promotion on prescription yield and sales uptake. The utilisation of

neural networks has been beneficial to 'learn' the relationships in the data based upon problem history, without knowledge about the nature of the relationship between the explanatory factors and the outcome. The principles of this work could therefore be easily extended to cover other sales prediction situations. It should, however, be borne in mind that the results of this study are not meant to supplant managers' decision making for specific product circumstances. Individual managers know better about their products and the particular context they face, so that information should greatly influence the final budget allocations. What this study does is to provide general guidance of what to do incrementally. In aggregate maybe that incremental dollar should go towards advertising spending. But it is not as literal as saying, 'Neural network is a magic wand'. It does not replace skilled managers, but rather gives them a powerful new tool to improve the job they are doing. The real challenge is determining how to apply the knowledge extracted through neural networks to business advantage.

While it is not too difficult to devise what seems to be an efficient approach for a company, a manager is not usually in the luxurious position of being able to implement that approach from scratch. Because of the realities deriving from organisation culture (the difficulty of organising a company-wide project while avoiding local jealousies and power struggles), and manpower and resource constraints, traditional methods still remain dominant as a favoured prediction tool in most companies. This approach may be powerful, but it only becomes usable within an organisation when it is shared. There needs to be a transition in executives' attitudes to a more open, mutually supporting and self-actualising culture.

ACKNOWLEDGMENTS

The authors wish to thank the Editor and the anonymous referees for many useful comments. Thanks are also extended to Mr Ken Masuo and Mr David Griggs for proofreading assistance.

References

- 1 Rosenthal, M. B., Berndt, E. R., Donohue, J. M. *et al.* (2003) 'Demand effects of recent changes in prescription drug promotion', The Henry J. Kaiser Family Foundation, Menlo Park, CA.
- 2 Corstjens, M. and Merrihue, J. (2003) 'Optimal marketing', *Harvard Business Review*, Vol. 10, pp. 114–121.
- 3 Altman, D. G. (1995) 'Practical statistics for medical research', Chapman & Hall Ltd, London.
- 4 Holmström, L. and Koistinen, P. (1992) 'Using additive noise in back-propagation training', *IEEE Transactions on Neural Networks*, Vol. 3, No. 1, pp. 24–38.
- 5 Alon, I., Qi, M. and Sadowski, R. J. (2001) 'Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods', *Journal of Retailing and Consumer Services*, Vol. 8, pp. 147–156.
- 6 Makridakis, S. and Wheelwright, S. C. (1989) 'Forecasting methods for management', 5th edn, Wiley, New York.
- 7 Akaike, H. (1973) 'Information theory and an extension of the maximum likelihood principle', in Petrov, B. N. and Csaki, F. (eds) 'Proceeding 2nd International Symposium on Information Theory', Akademia Kiado, Budapest, pp. 267–281.
- 8 Moody, J. (1992) 'The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems', in Moody, J., Hanson, S. J. and Lippmann, R. P. (eds) 'Advances in neural information processing systems', Morgan Kaufmann, San Mateo, CA, pp. 847–854.
- 9 Vapnik, V. (1995) 'The nature of statistical learning theory', Springer.
- 10 Stone, M. (1974) 'Cross-validated choice and assessment of statistical predictions', *Journal of the Royal Statistical Society B*, Vol. 36, pp. 111–147.
- 11 Schenker, B. and Agarwal, M. (1996) 'Cross-validated structure selection for neural networks', *Computers and Chemical Engineering*, Vol. 20, No. 2, pp. 175–186.
- 12 Castiglione, F. (2001) Forecasting price increments using an artificial neural network', *Advances in Complex Systems*, Vol. 4, No. 1, pp. 45–56.
- 13 Rosenthal *et al.* (2003) *op. cit.*
- 14 Healthy Skepticism Homepage (2003) 'Pharmaceutical giants over-invest in promotion', <http://www.healthyskepticism.org/reports/2003/0424.htm>, accessed 18th April, 2003.
- 15 Kallukaran, P. and Kagan, J. (1999) 'Data mining at IMS HEALTH: How we turned a mountain of data into a few information-rich molehills', in 'Proceedings of the 24th annual SAS users group international conference', Paper 127, <http://www2.sas.com/proceedings/sugi24/Dataware/p127-24.pdf>, accessed 11th December, 2003.
- 16 Herman, R. (2002) 'Direct-to-consumer advertising of prescription drugs grows rapidly but remains a small share of promotional efforts', Harvard School of Public Health Press Release, <http://www.hsph.harvard.edu/press/releases/press02132002.html>, accessed 13th February, 2002.
- 17 Berndt, E. R., Linda, B., David, R. *et al.* (1997) 'The roles of marketing, product quality and price competition in the growth and composition of the U.S. anti-ulcer drug industry', University of Chicago Press for the National Bureau of Economic Research, Chicago, pp. 277–322.
- 18 Rosenthal *et al.* (2003) *op. cit.*