

CTA vs. Disintegrated Chi-Square: Integrated vs. Piecemeal Analysis

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Experimental designs for which log-linear analysis is the recommended legacy (maximum-likelihood) methodology are usually inappropriately analyzed vis-à-vis series of chi-square analyses conducted on assorted subtables.¹ Disintegrated chi-square analysis is compared with CTA for an application relating physician support and desired smoking status to actual smoking behavior.

Data are drawn from a study assessing the relationship of physician encouragement and one's personal decision to stop smoking (attributes), and smoking behavior (class variable), for a sample of patients seen in two private medical practices.² Physician encouragement ("group") was a nominal indicator of whether the patient was seen by physician 1 who offered every patient who smoked encouragement to quit including medication and a pamphlet, or was seen by physician 0 who offered every patient who smoked nothing. Each patient's personal decision to stop (1) or to not stop (0) smoking was recorded, as was a follow-up measure of whether the patient stopped (1) or did not stop (0) smoking (see Table 1).

Chi-square analyses of assorted design subtables were reported, however the discussion focused on qualitative aspects of the research. The authors concluded: "it appears that the use of the physician's influence represents a potent way of altering smoking behavior" (p. 105).

CTA was used to analyze data in this application: smoking behavior was treated as the class variable, and group and personal decision were treated as categorical attributes.^{3,4} Only two models (the first of which was incomplete) were identified.

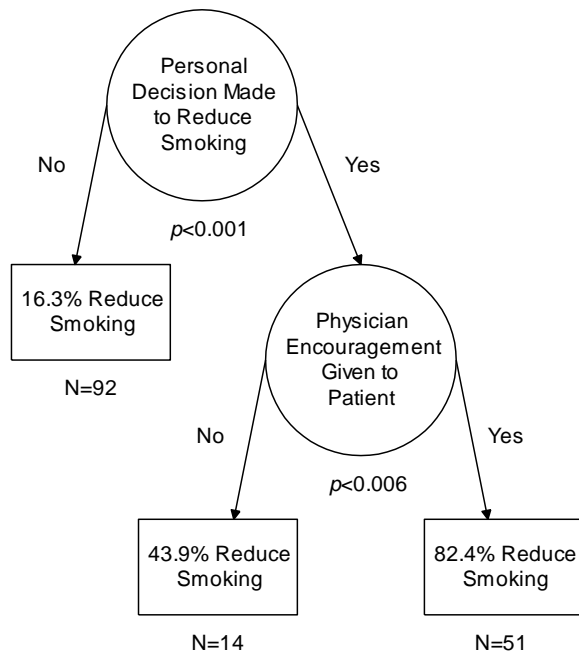
The first CTA model (Figure 1) was an incomplete hierarchically optimal (HO) CTA model that was fully-grown but not yet pruned to explicitly maximize ESS.³⁻⁶ As this model is unpruned it is unknown if it yields maximum possible normed predictive accuracy (ESS): the model is thus called a "suboptimal" HO-CTA model.^{3,4} As seen, this model included both of the attributes. The root (dominant) node was the patient's personal decision: of 92 people who decided they were not going to try to reduce smoking, 16.3% reported having reduced their smoking in the follow-up call. Of the remaining

Table 1: Smoking Study Data

<u>Group</u>	<u>Behavior</u>	<u>Stop</u>	<u>Continue</u>
Encourage	Stop	42	12
	Continue	9	58
Nothing	Stop	6	3
	Continue	8	19

people who decided they were going to attempt to reduce their smoking, 43.9% of 14 without physician support reported having reduced their smoking in the follow-up call. And, of the 51 people who intended to reduce smoking and received physician support, 82.4% reported reducing their smoking at follow-up.

Figure 1: Unpruned Three-Strata Suboptimal HO-CTA Model



This model refutes the conclusion made on the basis of the findings of disintegrated chi-square analyses, that physician influence is the primary agent in this application. Instead, the suboptimal HO-CTA model suggests physician influence is effective if the patient decides to act and the physician supports this decision.

Table 2 is the confusion matrix summarizing the predictive accuracy of this model: 9 in 10 patients who did not reduce smoking were correctly classified, and 2 in 3 patients who did reduce their smoking were correctly classified by the suboptimal HO-CTA model. This profile of predictive accuracy profile corresponds^{3,4} to a relatively strong effect (ESS=57.9): normed for

parsimony the model is $D=2.25$ equivalent effects distant from a theoretically ideal model achieving perfect accuracy and maximum parsimony in this application.⁴

Table 2: Confusion Matrix for Suboptimal HO-CTA Model

		Predicted Behavior		
		Same	Reduce	
Actual Behavior	Same	85	9	90.4%
	Reduce	21	42	66.7%

The second and final CTA model in Figure 2 was the suboptimal model (Figure 1) after being pruned to explicitly maximize ESS.

Figure 2: Pruned Two-Strata Globally-Optimal (GO) CTA Model

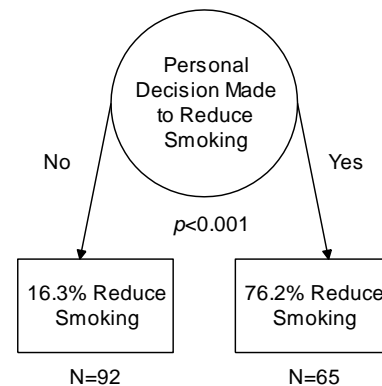


Table 3 is the confusion matrix summarizing the predictive accuracy of this model: 7 in 8 patients who did not reduce smoking were correctly classified, and 3 in 4 patients who did reduce their smoking were correctly classified by the GO-CTA model.

Table 3: Confusion Matrix for GO-CTA Model

		Predicted Behavior		
		Same	Reduce	
Actual Behavior	Same	77	17	81.9%
	Reduce	15	48	76.2%

This predictive accuracy profile corresponds to a relatively strong effect (ESS=58.1): the less complex, more parsimonious two-strata model yields greater normed predictive accuracy than the more complex, less parsimonious suboptimal three-strata model. Normed for parsimony the model is D=1.44 equivalent effects distant from a theoretically ideal model achieving perfect accuracy and maximum parsimony in this application. Because this model had the lowest D statistic of any (of the optimal) models identified for the sample, this is the GO model for this application.⁴

This model refutes the conclusion made based on findings of disintegrated chi-square analyses, that physician influence is the primary agent in this application. Indeed, the GO-CTA model demonstrates that physician influence adds nothing to the patient's decision, regardless of the nature of that decision.

References

¹Grimm LG, Yarnold PR (1995). *Reading and Understanding Multivariate Statistics*. Washington, D.C.: APA Books.

²Mausner B, Mausner JS, Rial WY (1967). The influence of a physician on the smoking behavior of his patients. In: Zagona SV (Ed.), *Studies and issues in smoking behavior*. Tucson, AZ: University of Arizona Press (pp. 103-106).

³Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC, APA Books.

⁴Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

⁵Yarnold PR, Bryant FB (2015). Obtaining a hierarchically optimal CTA model via UniODA software. *Optimal Data Analysis*, 4, 36-53.

⁶Yarnold PR, Soltysik RC (2010). Maximizing the accuracy of classification trees by optimal pruning. *Optimal Data Analysis*, 1, 23-29.

Author Notes

The study analyzed de-individualized data and was exempt from Institutional Review Board review. No conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646