

CONSUMER COMPREHENSION OF EFFICACY DATA IN FOUR EXPERIMENTAL OVER-THE-COUNTER LABEL CONDITIONS*

DOUGLAS B. GRISAFFE, PHD

Vice President, Research Sciences, Product Development and Management

SUSAN SHELLABARGER, MS

Vice President, Business Development, Healthcare Product Research

Walker Information, Indianapolis, Indiana

Recently marketed over-the-counter (OTC) drugs include efficacy data in their labeling/packaging. A mall intercept methodology was used with a hypothetical antacid product to investigate consumer comprehension of such efficacy data under four experimental label conditions. Consumers who were given no efficacy data were inaccurate in predicting the drug's effectiveness. People who were given efficacy data were much more accurate, but their accuracy depended upon which experimental label they received. A text presentation showed some advantage over graphical presentations. Other analyses point to a logical chain of effects from demographic and label characteristics, to comprehension and anticipation of efficacy, to purchase intentions. A model is presented which summarizes this chain of effects. This study has practical implications for consumers, the Food and Drug Administration (FDA), and manufacturers of OTC drug products. In fairness to the consumer, the ultimate link to purchase intention dictates label conditions that maximize comprehension and foster accurate anticipation of personal efficacy.

Key Words: OTC drug labeling; Efficacy data; Label comprehension; Demographic differences; Purchase intention

INTRODUCTION

IN AN ERA OF INCREASING health care costs, more responsibility for health maintenance is being shifted to individuals. Concurrently, powerful medications that once were available only by prescription (Rx) are now

being offered for over-the-counter purchase. For pharmaceutical companies to obtain approval by the FDA for such an "Rx-to-OTC switch," however, an approval process must be followed. Several steps in this process are likely to incorporate empirical data from carefully-conducted scientific research (1).

One place where it is critical to collect empirical data is in the area of label comprehension. Beyond just being able to read a label (legibility), consumers must be able to understand, interpret, and apply label information if they are to use an OTC product safely and correctly (1). In other words, com-

*Presented in part at the Institute for International Research "OTC Drug Market Conference," October 23-25, 1995, Philadelphia, Pennsylvania.

Reprint address: Doug Grisaffe, Walker Information, 3939 Priority Way South Drive, Indianapolis, IN 46240-0972.

In summary, it is important to consider both classes of influencing factors in the creation of labeling for OTC drugs. If the goal is to maximize comprehension, one must consider what does and does not affect it. Of the influencing factors, however, label characteristics are obviously most controllable by OTC drug manufacturers. This is because many decisions are in the hands of the manufacturers, subject to industry standards and regulations, concerning both the content and format of label information. For example, a drug interaction warning (content) could be presented as one long paragraph of text or as a series of bullet points (format).

Decisions about content and format should be made so as to maximize success of various objectives. One primary objective has been defined by the FDA. For example, Holt et al. (3) cite FDA regulations requiring OTC drug labeling that is readable and understandable by ordinary individuals, including individuals of low comprehension, under customary conditions of purchase and use. That should hold for any general consumer with access to the OTC product, not just the subpopulation with indications for use, so that all consumers are ultimately protected.

Presentation of Efficacy Data

Recently, there has been an effort to add efficacy information to the content of OTC labeling, with the data typically derived from "double blind" experiments.

As interesting and informative as efficacy studies are to the FDA and researchers, a number of key questions must be raised when considering the information from the perspective of the consumer:

- If consumers of OTC drugs are presented with efficacy data, are they of interest and benefit to them, or do they simply add to already-crowded labels?
- Can consumers understand the information?
- Are there ways to present efficacy data that maximize comprehension of it?
- Would such data help consumers make

more informed purchase and usage decisions?

- Would it help them to set realistic expectations about the effectiveness of the product?
- Would levels of understanding differ according to key consumer characteristics?

Despite a general absence of empirical answers to such questions, some people both inside and outside of the FDA hold the opinion that presentation of efficacy data would be helpful for consumers and should be provided as part of OTC drug labeling/packaging. Thus, several recently-approved OTC switch products have presented efficacy data as part of their labeling/packaging. As implied by the previous list of questions, however, it will be of little use to present the information if consumers cannot comprehend it. In fact, it might be that such information simply adds clutter to already-information-packed labels. Of course, questions about comprehension of efficacy data can be tested empirically. Thus, a study was designed to begin to address some research questions surrounding comprehension of efficacy information in OTC labels.

Overview of Methodology

In creating labeling for a hypothetical heartburn/acid reducing product, four variants formed the basis of the experimental conditions. All had a statement that clinical studies had shown the product to be significantly better than placebo tablets in bringing relief to symptoms. That was all that appeared on one of the labels, with no actual efficacy data presented. The other three labels all presented the same set of efficacy numbers, but in different presentation formats. One used a text format, one a graph format, and one an enhanced "graph plus" format.

Two types of subjects participated: those who reported use of OTC antacid products, and those who reported nonuse. These subjects were recruited in natural consumer settings (shopping malls) and randomly assigned to one of the four label conditions.

TABLE 1
Participant Demographic Profiles

Variable	Descriptive Statistics	
Age	Mean	Std. Dev.
	35.44	13.75
Gender	Frequency	Percent
Males	197	44.8
Females	243	55.2
Education Completed	Frequency	Percent
Grade School	1	.2
Some High School	37	8.4
Graduated High School	103	23.4
Some College	172	39.1
Graduated College	85	19.3
Graduate School	42	9.5
Ethnic Background	Frequency	Percent
Caucasian/White	291	66.7
Afro-American/Black	100	22.9
Asian	10	2.3
Hispanic	21	4.8
Native American	4	.9
Other	10	2.3
Annual Household Income	Frequency	Percent
Less than \$20,000	91	21.8
\$20,000-39,999	145	34.7
\$40,000-59,999	100	23.9
\$60,000 or more	82	19.6

told that they could refer to the label as often as they liked while filling out the questionnaire. If a participant had a question about the label or the questionnaire, a "pat" answer was given indicating that no assistance could be provided. Those with questions were encouraged simply to do the best they could, or to write in a "don't know" response. Participants received no compensation but were thanked upon completion.

LABELS

Two recent Rx-to-OTC switch products already existed in this product category (Pepcid AC Acid Controller and Tagamet HB). The features of their labeling were carefully studied. Those features became the basis for the study materials. The authors did not, however, wish to use either existing product label in its exact form because users of those brands would then have had prior experience with the materials. Further, brand is such a

powerful information cue that leaving actual brand names on the materials would have produced uncontrolled effects based on the participant's past experiences. Thus, the dimensions on which the two existing labels differed were identified, and then those features were built into the conditions of the study.

Four labels were constructed for the hypothetical product in this category. These are shown in Figure 1, Figure 2, Figure 3, and Figure 4. The product was called Nonprescription H-A-R[™] Tablets: Heartburn Acid Reducer. The four different H-A-R labels served as the basis for the experimental design. The physical creation of this labeling was done by a professional graphic designer who had experience in OTC and health care-related package designs.

The four H-A-R labels defined the conditions to which participants were randomly assigned. The first label was used as a control condition. No explicit efficacy data were pre-

Non-prescription H-A-R[™] Heartburn Acid Reducer

Active Ingredient:

The active ingredient in Non-Prescription H-A-R[™] is a medicine doctors have prescribed safely many times, for many years to reduce stomach acid that can cause heartburn.

Uses:

Use Non-Prescription H-A-R[™] to treat heartburn, acid indigestion, sour stomach.

Directions:

- To treat symptoms, take 1 tablet with water.
- Non-Prescription H-A-R[™] can be used up to twice daily (up to 2 tablets in 24 hours).
- This product should not be given to children under 12 years old unless directed by a doctor.

Results Of Clinical Studies:

In clinical studies, Non-Prescription H-A-R[™] was significantly better than placebo tablets in relieving and preventing heartburn symptoms.

Percent of heartburn episodes relieved:		Percent of patients with prevention or reduction of heartburn symptoms:	
H-A-R [™] Tablet = 72%	H-A-R [™] Tablet = 67%	H-A-R [™] Tablet = 74%	
Placebo Tablet = 46%	Placebo Tablet = 49%	Placebo Tablet = 55%	
Study A	Study B	Study C	

Warnings:

- Do not take the maximum daily dosage for more than 2 weeks continuously except under the advice and supervision of a doctor.
- As with any drug, if you are pregnant or nursing a baby, seek the advice of a health professional before using this product.
- If you have trouble swallowing, or persistent abdominal pain, see your doctor promptly. You may have a serious condition that may need different treatment.
- Keep this and all drugs out of the reach of children.
- In case of accidental overdose, seek professional assistance or contact a poison control center immediately.

FIGURE 2. Text label condition.

Non-prescription H-A-R[™] Heartburn Acid Reducer

Active Ingredient:

The active ingredient in Non-Prescription H-A-R[™] is a medicine doctors have prescribed safely many times, for many years to reduce stomach acid that can cause heartburn.

Uses:

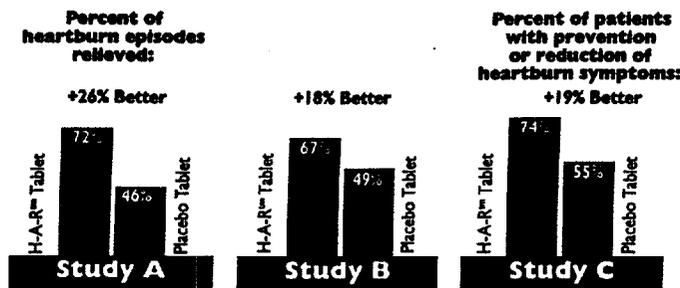
Use Non-Prescription H-A-R[™] to treat heartburn, acid indigestion, sour stomach.

Directions:

- To treat symptoms, take 1 tablet with water.
- Non-Prescription H-A-R[™] can be used up to twice daily (up to 2 tablets in 24 hours).
- This product should not be given to children under 12 years old unless directed by a doctor.

Results Of Clinical Studies:

In clinical studies, Non-Prescription H-A-R[™] was significantly better than placebo tablets in relieving and preventing heartburn symptoms.



Warnings:

- Do not take the maximum daily dosage for more than 2 weeks continuously except under the advice and supervision of a doctor.
- As with any drug, if you are pregnant or nursing a baby, seek the advice of a health professional before using this product.
- If you have trouble swallowing, or persistent abdominal pain, see your doctor promptly. You may have a serious condition that may need different treatment.
- Keep this and all drugs out of the reach of children.
- In case of accidental overdose, seek professional assistance or contact a poison control center immediately.

FIGURE 4. Graph plus label condition.

Arguably, it would have been more straightforward to present efficacy data from only one clinical study. Then there would be a single standard of correct and incorrect comprehension of that information. The presentation of data from multiple studies, however, was based on actual labeling from the two existing OTC products in this category. Each of those existing products present data from multiple studies of efficacy, leading the authors to do the same for this research. In fact, the numbers from the real labels were worked with to arrive at realistic efficacy numbers for the hypothetical product, H-A-R.

Assuming that participants perform some type of mental averaging, it is simple enough to come up with a standard of correctness for the three fill-in-the-blank items about the effectiveness of H-A-R, the placebo, and the difference in their efficacies. As such, a standard of correctness was calculated by averaging the multiple-study data presented in the materials. Average efficacy of the product was $(72 + 67 + 74)/3 = 71\%$. Average "efficacy" of the placebo was $(46 + 49 + 55)/3 = 50\%$. The difference in these averages, 21%, served as the standard of correctness for the question about difference in placebo and H-A-R effectiveness.

Obviously, many respondents will not compute exact arithmetic averages when responding to the questionnaire items. Therefore, it would be unreasonable to score these three items dichotomously as correct or incorrect. An approach was taken that was thought to be more reasonable. The exact averages described previously were used as a standard against which each respondent's data could be compared. A quantity known as the absolute deviation was computed (14). This is simply the absolute value of the difference between the response and the standard. Thus, if a respondent filled in 75% for H-A-R efficacy, 40% for placebo efficacy, and 35% for the difference, his/her absolute deviation measures would be $|75 - 71| = 4$, $|40 - 50| = 10$, and $|35 - 21| = 14$, respectively.

For other analyses in the study, a categori-

cal version of the respondent's evaluation of H-A-R efficacy was created. Respondents were classified into one of three categories: correct estimate, underestimate, and overestimate. A person was placed in the correct estimate category if his/her estimate of H-A-R effectiveness was within plus or minus five percentage points of the range of effectiveness indicated by the label information. Any estimates lower than that range were categorized as underestimates. Any estimates above that range were categorized as overestimates.

Placebo Item. Participants were asked to provide a description/definition for the concept of a placebo tablet. These open-ended responses were then scored by two graders. Four codes were used. Any respondent could have given a correct definition, an incorrect definition, left the item blank, or written in a "don't know" response. Blanks and "don't knows" were straightforward to code. For the correct and incorrect codes, a systematic grading procedure was followed.

Five of the 440 questionnaires were selected at random. The first author and the two graders had an open discussion about the responses on these five questionnaires. Next, 25 more questionnaires were selected at random. They were divided among the graders, and scored on a separate score sheet with one of the four codes. The graders then switched their questionnaires and continued scoring. In the end, both graders had coded all 25 questionnaires. Interrater reliability in coding was at an acceptable level with 23 out of 25 agreements (92%). The two disagreements were discussed and consensus was achieved.

From the random selection of 30 questionnaires in the reliability exercise, a list of correct responses and a list of incorrect responses (see Appendix) was generated. These were used for comparison purposes for all further coding. To score the remaining 410 questionnaires, they were divided among the two graders. The graders were instructed to set aside any responses that they did not know how to score. Uncertainty about grad-

bivariate correlation between the two measures was used to assess this association. A cross tabulation approach was also applied. It was hypothesized that anticipated personal effectiveness would show a strong linear relationship with purchase intention. That link might be the final outcome in a chain of effects that start with comprehension of label information.

RESULTS

Comprehension of Control Items

Comprehension was dichotomous for the first three questionnaire items. To test if this depended upon the two dimensions of the design, frequencies in contingency tables were analyzed using log linear analysis. Correct/not correct, user/nonuser, and label conditions one through four formed the cells of the $2 \times 2 \times 4$ table for each control item. A log-linear model with a specific set of model terms was fitted to the data.

For each control item analyzed, the model applied has been referred to by Kennedy as the "null logit model" (15). This model leaves out all terms representing the dependency of correct/not correct on the other categorical factors (user type, label condition, and their interaction). If this model fits (ie, sufficiently reproduces all cell frequencies in the full table), then it can be concluded that correct responses to the item do not depend on the respondent's user status, label condition, or the interaction of user status and label condition (16). Results of these tests are shown in Table 2.

It is clear that comprehension of these control items did not depend significantly on

the factors of the main design. Thus, it is legitimate to report comprehension of them in total. Correct responses to the control items were given by 94.3%, 96.4%, and 90.0% of the sample for items Q1, Q2, and Q3, respectively.

Comprehension of Efficacy Data by Experimental Condition

As hypothesized, the correlations among the computed comprehension measures for the efficacy data were all significant at $p < .001$ (.551 [Q4&Q5], .258 [Q4&Q6], .398 [Q5 & Q6]). Because of the fact that these dependent measures were correlated and that it was suspected that they would all vary according to the dimensions of the 2×4 experimental design, the measures were analyzed using MANOVA. Besides handling correlated dependent measures, an additional side benefit of MANOVA is control over experiment-wise Type I error. The benefits gained with MANOVA, however, come at the expense of stricter assumptions about the data. Statistical assumptions call for multivariate normality as well as homogeneity of variance-covariance (17).

While the method of computing scores as absolute deviations from a standard of correctness was logical in operationalizing comprehension, it also produced highly skewed data. A large proportion of people had values close to the standard of correctness. Frequencies of respondents tended to trail off going farther and farther away from the standard. Therefore, to more closely meet statistical assumptions, the data were log-transformed, as recommended by Tabachnik and Fidell (17), before applying MANOVA. Note that

TABLE 2
Fit of Null Logit Models for Control Items

Control item	Likelihood Ratio		
	Chi Square	DF	P-value
Q1. Consult healthcare if pregnant	8.40	7	.298
Q2. H-A-R designed for headaches	9.34	7	.229
Q3. Number of allowable doses per day	9.21	7	.238

TABLE 4
Untransformed Efficacy Comprehension by Label Condition

	Control Label	Text	Graph	Graph Plus
H-A-R Effectiveness				
Mean	21.93	7.79	9.11	14.62
Std. Dev.	17.89	14.56	15.74	19.91
Placebo Effectiveness				
Mean	19.06	8.42	9.82	12.18
Std. Dev.	16.61	11.74	12.59	14.97
Difference in Effectiveness				
Mean	20.64	13.31	15.03	11.19
Std. Dev.	18.17	14.98	17.12	16.43

of text over graphs, and the regular graph over the graph plus for comprehension of drug efficacy data. The only other additional piece of information comes with respect to comprehension of the difference in drug and placebo effectiveness. It appears that the difference is understood most accurately when explicitly shown in the graph plus condition. The pattern of means in Table 4 highlights these substantively plausible findings.

Directionality of Inaccurate Comprehension

Initial exploratory work led to the use of absolute deviations from a standard of correctness in the previous analyses. Without taking the absolute values, an arithmetic problem arises when evaluating means because deviations from the standard can occur in one of two ways. Some people underestimate effectiveness while others overestimate effectiveness. These "positive" and "negative" errors tend to cancel each other out. Taking the absolute value solves this problem

and provides a good view of absolute correctness. Information about directionality of inaccurate estimates, however, is lost.

In a follow-up analysis, the deviations from the standard of correctness were recorded into one of three categories: correct estimate, underestimate, and overestimate. Based on the previous analysis, results are presented only for perceived H-A-R effectiveness. A person was placed in the correct estimate category if his/her estimate of H-A-R effectiveness was within plus or minus five percentage points of the range of effectiveness indicated by the label information. Any estimates lower than that range were categorized as underestimates. Any estimates above the range were categorized as overestimates. The result was a three-category dependent variable, leading again to the use of log-linear modeling. Results are shown in Table 6.

In this case, the null logit model does not fit the data. Adding a term capturing H-A-R comprehension's dependency on user type does not work to make the model fit better. The only thing that needs to be added to the

TABLE 5
Tests of Helmert Contrasts for Efficacy Comprehension Measures

	Transformed Q4		Transformed Q5		Transformed Q6	
	T	P-value	T	P-value	T	P-value
Contrast 1	10.89	.000	3.53	.001	4.85	.000
Contrast 2	-2.12	.035	-1.33	.184	.57	.573
Contrast 3	-2.07	.040	-.63	.531	2.35	.019

concept of a placebo. Thus, there must be no relationship between comprehension of the placebo concept and the cells of the experimental design. That should not be a concern if the randomization was done correctly, but it is an important alternative explanation to be ruled out.

A log-linear approach was used to test the relationship between understanding of the placebo concept and the cells of the experimental design. The null logit model fit the data, with a Likelihood Ratio Chi Square = 4.08, $df = 7$, $p = .770$. Thus, general comprehension of the placebo concept did not depend upon any factors in the design.

People who understand the placebo concept also might do better in general comprehension given that the comparison of drug to placebo would make so much more sense to them. Thus, tests of efficacy comprehension were conducted comparing those who did and did not understand the placebo concept. This was done in the context of the 2×4 MANOVA design so as to control for the other known influences on the dependent variables. In other words, the 2×4 design became a $2 \times 4 \times 2$ design by adding in the dichotomous placebo comprehension measure as another independent variable.

Using the transformed absolute deviation dependent variables, MANOVA revealed a significant main effect for placebo understanding, $F(3,354) = 5.75$, $p = .001$. No interactions were significant and the results for the original design factors did not change. Single dependent variable tests with $df(1,356)$ were significant for comprehension of H-A-R effectiveness $F = 15.01$, $p < .001$, and for comprehension of placebo effectiveness, $F = 5.66$, $p < .02$. The effect on comprehension of the difference of H-A-R and placebo was marginal, $F = 2.79$, $p < .10$. Means on the untransformed variables are presented in Table 8.

These are not surprising results. It might be expected that people who understand the concept of placebo would be more likely to perform well in comprehending clinical efficacy data.

Comprehension in Relation to Demographics. The final analyses in relation to comprehension explored demographics. A series of bivariate analyses were conducted to explore how various demographic characteristics did or did not relate to the primary comprehension measures of H-A-R efficacy. Demographics were explored in relation to three of the previous measures of comprehension. Two measures were categorical: placebo definition (correct or incorrect), and categorical H-A-R estimate (underestimate, correct estimate, or overestimate). For these measures, Chi-square tests of independence were conducted with each demographic variable. The third comprehension measure, absolute deviation from the standard of correctness, was treated as a dependent variable and one-way analysis of variance was conducted with each demographic serving as an independent variable.

Note that no correction was applied for the inflation of Type I error rates that comes with repeated statistical testing. Even though multiple tests were being conducted, they were totally exploratory and descriptive. The purpose of these analyses was simply to gain a better understanding of individual factors that might relate to comprehension. Results are presented in Table 9. It should also be noted here that the effects of these demographic factors did not account for the previous findings. When demographic factors were statistically controlled in a second pass at the original 2×4 MANOVA analysis, substantive results did not change.

Perceived General Efficacy in Relation to Expected Personal Efficacy

A Path Analytic Approach. Having explored comprehension in relation to the experimental design and various background characteristics, attention was turned to the personal interpretation and application of the label information. Comprehension of general label information is not a sufficient stopping point for OTC drugs. Consumers must be able not only to understand the information on the

label, but also to interpret and apply it relative to their own conditions. Thus, comprehension of general efficacy data was explored to see how it might or might not affect consumer expectations about effectiveness in their own personal application.

Respondents had been asked to use a five-point scale to rate how effective they thought H-A-R would be for them if they tried the product. This was done to test how this measure of personal efficacy related to their general estimate of H-A-R efficacy. In other words, do respondents make the leap from general comprehension to personal application?

Having already discovered that label conditions affect general perceptions of product efficacy, it was felt to be important to conceptualize the general-to-personal test in the context of the experimental design. To allow for the possibility that label and user conditions might have a direct effect on perceptions of personal effectiveness, as well as an indirect effect through perceptions of general effectiveness, a path analytic framework was conceptualized to represent all logical effects. This is presented in Figure 5. While all paths were analyzed, note that the most interest was in testing the path from general effectiveness to personal effectiveness, controlling for any direct effects from the label and user factors.

Before applying regression, the measures

of general effectiveness had to be reconsidered. Absolute deviations from correctness were good for testing general accuracy using means. But here, at a minimum, whether deviations from correct were overestimates or underestimates had to be represented. In other words, for the purposes of the current test, it was critical to retain information about the true magnitude of estimated effectiveness of H-A-R. If a respondent read a label and gave a low estimate of general H-A-R effectiveness, then it is likely that he/she would believe it to be less effective in his/her own personal application. Likewise, a respondent who read the label and estimated H-A-R to have high general effectiveness might have high personal expectations of effectiveness. Thus, the magnitude of perceived general effectiveness had to be retained to test for the linear general-to-personal link addressed in this section. Therefore, the raw estimates of effectiveness were used without any deviation calculations, or absolute values, or log transformations.

The paths represented in Figure 5 were estimated with ordinary least squares regression. The dependent variable was the five-point measure of perceived personal effectiveness. Among the independent variables were each of the three raw estimates of general effectiveness. A set of dummy variables was also constructed to represent all of the experimental design factors (user type, label

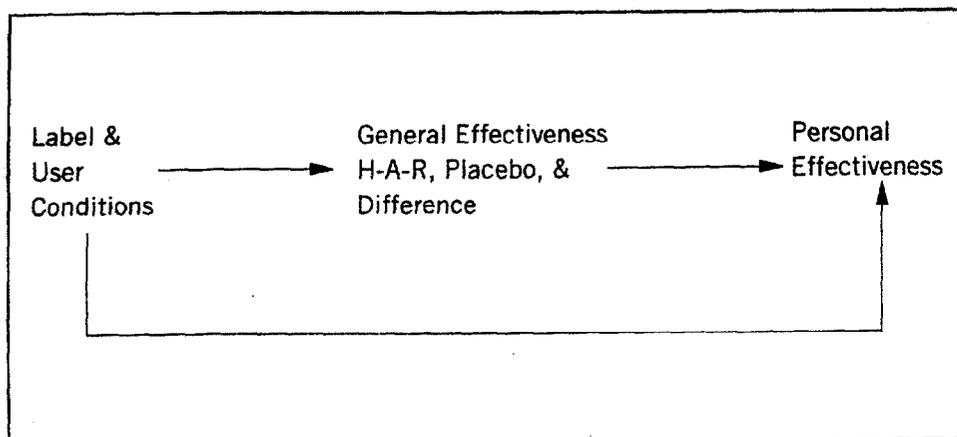


FIGURE 5. Path analytic framework for general-to-personal efficacy test.

TABLE 11
ANOVA Tests of Variables on Personal Effectiveness

Source of Variation	SS	DF	MS	F	Sig of F
User Type (U)	2.54	1	2.54	3.68	.056
Label Version (L)	5.09	3	1.70	2.46	.062
General Effectiveness (G)	5.95	2	2.98	4.32	.014
U x L	3.27	3	1.09	1.58	.193
U x G	2.34	2	1.17	1.70	.184
L x G	6.76	6	1.13	1.64	.136
U x L x G	6.71	6	1.12	1.62	.139
Error	249.38	362	.69		

factor that had an unambiguous significant effect on perceived personal effectiveness. Means by the categorical version of general effectiveness are presented in Table 12.

As a follow-up test of these means, a planned contrast was used to represent and test the expected linear trend, that is, the expectation of higher levels of perceived general effectiveness relating to higher levels of personal effectiveness was tested. The follow-up contrast for the linear effect was statistically significant: $F(1,383) = 8.18$ $p < .005$.

Thus, when using the measure of general effectiveness as either a continuous variable or as a categorical variable, a significant relationship between perceived general effectiveness and perceived personal effectiveness was found, even after controlling for the other known influences in the design.

Links Between Perceived Personal Efficacy and Purchase Intention

A purchase intent question on a five-point likelihood scale was included. Respondents

were asked their likelihood to purchase H-A-R if they were experiencing heartburn, acid indigestion, or sour stomach and had decided to buy a medicine to treat it. It was expected that this likelihood to purchase H-A-R should increase linearly with the perception of personal effectiveness. Indeed, this was the case. The bivariate correlation between the two measures with 416 degrees of freedom was $r = .544$, $p < .001$. Further, this correlation was not moderated by the factors of the design. A test of homogeneity of correlations within the eight cells of the experimental design was nonsignificant. Thus, purchase intention, not surprisingly, was strongly related to perceived personal effectiveness. A graphical view of this relationship is presented in Figure 6.

DISCUSSION

The study findings can be summarized as follows:

1. Most consumers (82.3% in the sample) who receive no efficacy information can-

TABLE 12
Personal Effectiveness by Categorical General Effectiveness

	General Effectiveness Underestimate	General Effectiveness Correct Estimate	General Effectiveness Overestimate
Personal Effectiveness			
Mean	2.89	2.69	2.48
Std. Dev.	.84	.82	1.04
n	117	217	52

background demographic characteristics (except race) do not seem to be strongly related to general label-based efficacy comprehension, they are highly related to comprehension of the placebo concept. Those who understand the placebo concept, in turn, have significantly higher comprehension of label-based efficacy data. Comprehension of efficacy data is also influenced by presentation format. Once all these effects express themselves on people's general perceptions of product efficacy, those general perceptions influence people's anticipation of effectiveness in personal application. It would appear that label presentation format also has some marginal direct influence on anticipated personal effectiveness. Finally, once the expectation of personal effectiveness is established, it heavily determines consumers' behavioral purchase intentions.

The chain of effects implied by these results is shown in Figure 7. It is believed that this diagram accurately captures the general essence of the findings. It should be noted, however, that the idea of such a chain of effects is certainly not new. Other authors have postulated similar models (4,7,9). The diagram has been used as an *ex post facto* tool to represent the findings rather than as an *a priori* conceptual framework to design the research.

Specific elements of the results relate to

findings in previously published research. For example, Sansgiry and Cady (20) found that explicit text statements of information outperformed pictorial representations. Others, however, have cited research to the contrary (5). Since there are many ways to present graphical information, it is likely that graphical methods vary in effectiveness, with some designs outperforming others. Perhaps another version of graphic format in the study would have produced different results. Discovery of optimally effective graphic presentations is something that could be explored in future empirical research.

Also related to prior research is the finding that demographics impact comprehension only indirectly through the mediating comprehension of the placebo concept. Holt et al. (3) explained their lack of relationship between demographics and comprehension by postulating that interpretation of labels is somewhat universal in the United States as a by-product of general enculturation with the American health care system. As described in the introduction, however, the mixed findings in the literature between demographics and label comprehension might diminish the strength of that explanation.

Perhaps mediation explains the mix of results. In other words, the effect sometimes might be only indirect through mediating variables. At other times, perhaps the effect is

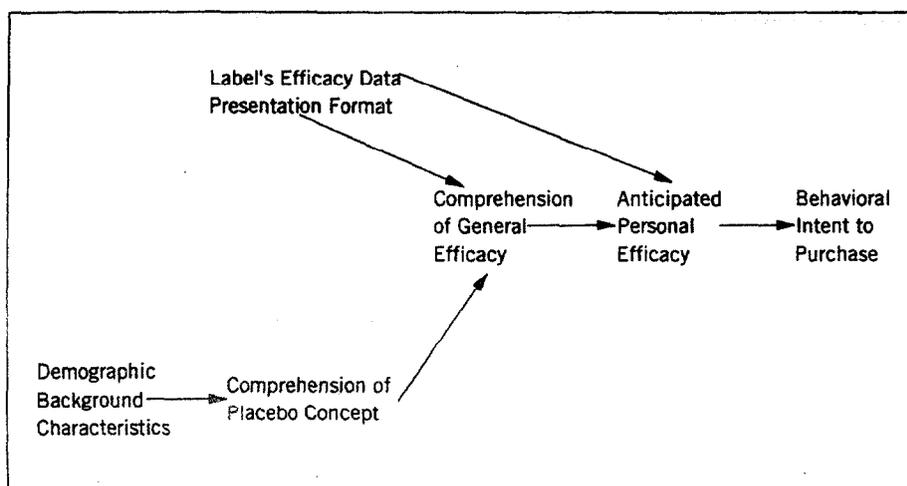


FIGURE 7. Conceptual summary of results.

14. Elzey F. *Introductory Statistics: A Microcomputer Approach*. Monterey, CA: Brooks/Cole Publishing Company; 1985.
 15. Kennedy JJ. *Analyzing Qualitative Data*. New York, NY: Praeger Publishers; 1982.
 16. Agresti A. *Categorical Data Analysis*. New York, NY: John Wiley & Sons; 1990.
 17. Tabachnik BG, Fidell LS. *Using Multivariate Statistics*. New York, NY: Harper & Row; 1983.
 18. Norusis MJ. *SPSS for Windows Advanced Statistics Release 5*. Chicago, IL: SPSS, Inc.; 1992.
 19. SPSS Inc. *SPSS-X User's Guide*. Chicago: SPSS Inc.; 1988.
 20. Sansgiry SS, Cady PS. The effect of label content and placement on consumers' understanding of otc product label information. *J Pharma Market Manage*. 1995;9(3):55-67.
2. No medicinal value
 3. Something neutral
 4. Does not affect person (but you think it does)
 5. Fake pill
 6. Used to make people think they're taking something
 7. Substitute tablet/pill

Sample Placebo Definitions Coded as Incorrect

1. Heartburn/antacid/ulcer medication
2. Less effective than brand
3. Not the same ingredients
4. Alternative to H-A-R
5. Generic/nonbrand
6. Tablet to relieve pain
7. Prescription Drug

APPENDIX

Sample Placebo Definitions Coded as Correct

1. Sugar pill