# Case Study: Direct Mail Subscriptions Tests

*Fast, low-cost creative and price tests*
*leading to a 70% jump in response*

## Introduction

Magazine publishers depend on direct mail to grow their subscriber base. But tight budgets and unresponsive consumers create a serious challenge—how to grow sales while mailing less.

One vice president of consumer marketing decided it was time to break the downward spiral. With a limited budget, she searched for a way to improve marketing ROI, but with greater speed, efficiency, and pinpoint accuracy. After talking with a few other industry leaders, she decided to try scientific testing.

In only two direct mail tests, this marketing team discovered more than they had learned in years of split-run testing. More importantly, they zeroed in on a few key, low-cost changes that together led to a 70% jump in response:

- **One test of 11 creative elements**—6 envelope, plus insert, letter, and order form changes—pinpointed 6 significant elements for a 55% jump in response (using 1/7$^{th}$ the usual sample size)

- **One test of 5 price and offer elements** showed how to increase response *and* profitability through a surprising two-way interaction (that split-run tests would never uncover)

## Step #1: Creative Freedom

The marketing vice president found a test expert, selected an important direct mail program to focus on, and scheduled the kick-off meeting.

In the first meeting, the consultant introduced the new approach and led a brainstorming session. Starting with their control mail package—the mailing that had a high and consistent response rate over the years—the team brainstormed 63 changes that might increase response and profitability. Ideas included changes to the outer envelope, order form, letter, inserts, price and offer.

## Step #2: Scientific Discipline

The next step was to concentrate the power of their ideas. With guidance, the team narrowed the list to 11 direct mail elements, focusing on bold, independent, and actionable changes. The team then chose two levels to test for each. The "minus" level was usually the control. The "plus" level was a change that someone thought would increase response rate or reduce costs.

| | Test Elements | (-) Control | (+) New Idea |
|---|---|---|---|
| A | Envelope color scheme | Control | New color palette |
| B | Background graphic on envelope | No | Yes |
| C | Second window | No | Yes, showing magazine cover |
| D | Expiration date display | Control | Show through envelope window |
| E | Teaser on envelope | Value message | Benefits message |
| F | Starburst on envelope back | No | Yes |
| G | Free gift buckslip | Yes | No |
| H | Lift note | Yes | No |
| I | Order form layout | Control | New layout |
| J | Letter length | Long | Short |
| K | Letter copy style | Control | New copy |

**A: Envelope color scheme** – The creative group came up with new, brighter colors to use for the fonts and graphics. This new color palette was tested against the standard color scheme on the envelope.

**B: Background graphic on envelope** – The team wanted to add a product-related picture to catch the attention of their target readers. They decided to test a background graphic in a light shade on the right side of the main address window.

**C: Second window** – To provide a stronger enticement to open the envelope, they wanted to test a second window on the left side, showing the full-color magazine cover on the order form inside.

**D: Expiration date display** – As a way to encourage people to respond immediately, the team wanted the expiration date to stand out more clearly, so they created a larger envelope window that shows the expiration date above the recipient's name.

**E: Teaser on envelope** – The teaser—a short phrase on the outside of the envelope—currently focused on the low cost of the subscription. One copywriter thought that a benefits-focused message would be a stronger incentive to subscribe.

**F: Starburst on envelope back** – The back of the control envelope was fairly plain. The team decided to test a bright starburst with a brief call-to-action message, supporting the main message and graphics on the front and inside.

**G: Free gift buckslip** – The offer included a free gift with each paid subscription. The gift was mentioned in the letter and on the order form and described in detail on a full-color separate sheet of paper in the mailing (a "buckslip"). One marketing manager thought that eliminating the buckslip would have little impact on response and reduce printing costs.

**H: Lift note** - Control mailings included a small note from the editor explaining the special benefits of the magazine. This "lift note" did not cost as much as the buckslip, but the team wanted to know if it really helped response. They considered testing a different version of the lift note, but decided instead to test eliminating it.

**I: Order form layout** – The creative team often rearranged the information, words, and graphics on the order form in each mailing. With no solid data showing which layout was best, they selected two different versions to test.

**J: Letter length** – Most mailings were sent with a separate letter explaining the magazine and the offer. The team believed that a long letter gave prospects more information, but they wondered if a short letter may help speed the decision process.

**K: Letter copy style** – Along with length (J), the team wanted to test a new message and positioning. With two different lengths, the copywriter created a long and short version with the new copy style, plus a short version of the control letter.

## Step #3: Statistical Power

Using scientific testing techniques, the consultant created a 12-recipe test design to simultaneously test all 11 elements in one mail drop. Each of the 12 versions included a unique combination of plus and minus levels of all 11 elements. Whereas each split-run test cell gives one data point on one variable, each scientific test recipe provides a new piece of information about every variable in the test.

With the same number of versions as the team would need for split-run testing (11 test cells + control), the scientific test design, below, offers important advantages:

- **Sample size** can be the same as for a simple one-variable test (an 86% reduction vs. A/B splits)
- **Main effects** are more accurate and robust (clearly quantify the impact of each element)
- **Interactions** between elements can be analyzed (showing how effects change in combination)
- **Optimal mail package** can be created by implementing the new ideas that help, avoiding changes that hurt, and selecting the cheaper alternative for elements that have no impact.

| | Envelope color scheme | Background graphic on envelope | Second window | Expiration date display | Teaser on envelope | Starburst on envelope back | Free gift buckslip | Lift note | Order form layout | Letter length | Letter copy style | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Recipes** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **Response** |
| 1 | + | - | + | - | - | - | + | + | + | - | + | 4.79% |
| 2 | - | + | - | - | - | + | + | + | - | + | + | 3.64% |
| 3 | + | + | - | + | - | - | - | + | + | + | - | 3.84% |
| 4 | + | - | - | - | + | + | + | - | + | + | - | 4.32% |
| 5 | - | + | + | - | + | - | - | - | + | + | + | 5.56% |
| 6 | - | - | - | + | + | + | - | + | + | - | + | 4.61% |
| 7 | + | - | + | + | - | + | - | - | - | + | + | 5.65% |
| 8 | - | - | + | + | + | - | + | + | - | + | - | 5.53% |
| 9 | + | + | - | + | + | - | + | - | - | - | + | 4.50% |
| 10 | - | + | + | + | - | + | + | - | + | - | - | 5.31% |
| 11 | + | + | + | - | + | + | - | + | - | - | - | 5.85% |
| 12 | - | - | - | - | - | - | - | - | - | - | - | 4.01% |

The key metric for the test was response rate—the percentage of recipients who mailed in the subscription order form. Overall sample size for test was 201,000 names, divided equally among all 12 test recipes (16,750 per recipe). With a 4.0% average response rate for the control, this sample size meant that the test had a good chance of uncovering elements that shifted response by about 6% (from 4.0% to 4.24%) and a 50-50 chance of uncovering effects as small as 4.3%.

The consultant gave the creative team 'recipe sheets" listing the combination of elements in each version, based on the above matrix. Since the test elements were clearly defined before the creative group got to work on the recipes, all 12 versions of the mailing added just three days to the creative schedule. Each recipe was labeled with a different keycode and randomly assigned 16,750 names.

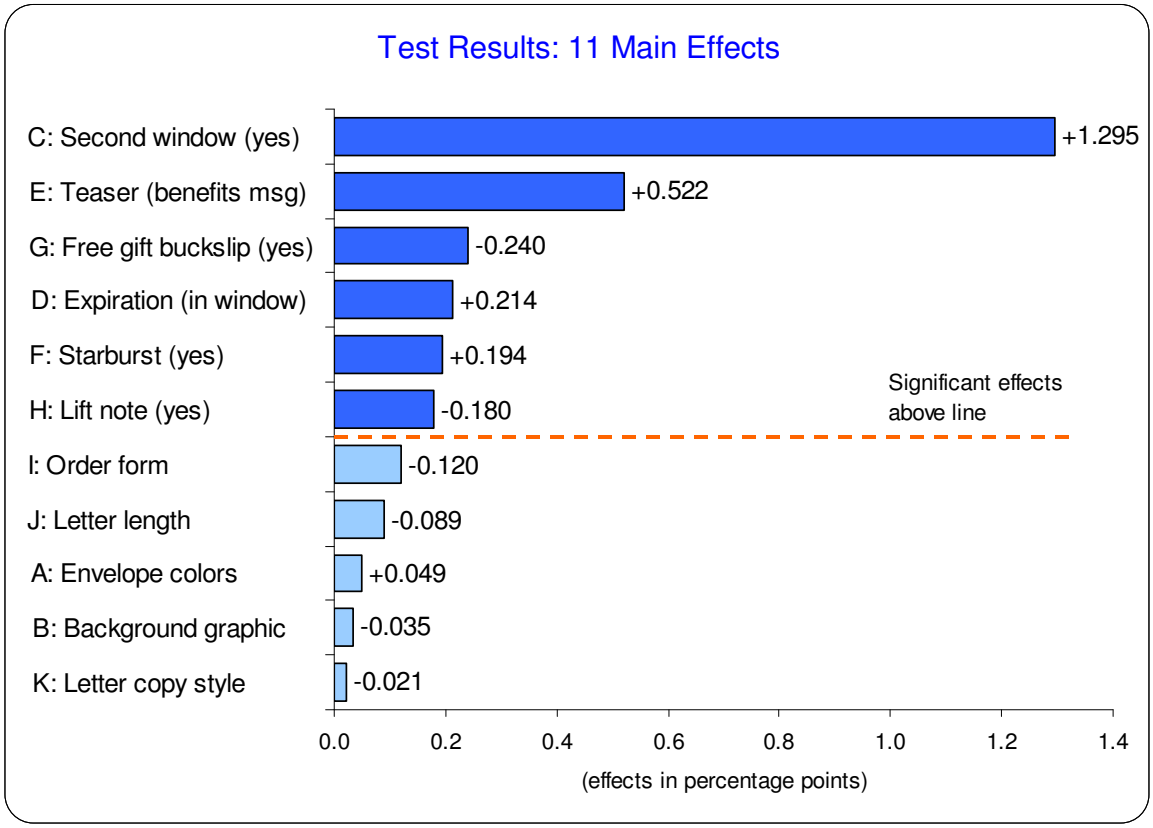
Eleven new outer envelopes were created (for elements A-F), one new order form (I), and three new letters (J, K). Recipe #12 is simply the control package. Each other recipe had about half the elements set at the control (-) level and half set at the new (+) level.

For example, recipe #1 required:

- An envelope with the new color scheme (A+), no background graphic (B-), a second window (C+), the control address window (D-), the control teaser (E-), and no starburst on back (F-)
- No buckslip (G+) or lift note (H+)
- New order form (I+)
- Long letter (J-) with the new copy style (K+)

## Step #4: Marketing Insights

All 12 recipes were produced, double-checked, and mailed on the same day. Following the team's standard response curves, results were analyzed after four weeks. Six elements were statistically significant, increasing response more than 50%. The 11 effects are summarized in the bar chart, below.



This simple chart gave the marketing team an easy way to see the size of each main effect, the optimal level of each, and which effects are statistically significant. In the chart, effects are ordered from largest (at top) to smallest (at bottom). The element name is on the left (with the optimal level in parentheses) and the effect is shown on the right as the length of each bar and the number at the end. The sign of each effect shows which level is better: with a positive effect, the "+" level will increase response and with a negative effect, the "–" level will increase response.

For example, the largest effect was the second window (C), with an effect of +1.295. Adding the second window on the envelope to show the magazine cover increases response 1.295 percentage points, from 4.0% to about 5.3% (a 32.4% increase in the number of responses). Also, with 9,649 total orders, overall response rate for the test was about 4.8% versus 4.0% for the control (recipe #12).

## Significant Effects

With this straightforward summary of results, the marketing team could analyze various combinations and the profitability of each. Adding the four positive effects (C, E, D, and F) to the control response rate gives a predicted response rate of 6.225%, a 55.6% jump in response over the control! The six significant effects include:

**C+: Second window**
The team was suprised at the large impact of this fairly simple change. The second window added little cost and gave an impressive 32.4% jump in response. They also realized the power of a full-color magazine cover and the importance of building interest before people even open the envelope. This change made so much sense after-the-fact, but the team had never thought to test it before.

**E+: New teaser (benefits message)**
The teaser with a message about the benefits of the magazine increased response 0.522 percentage points (13%) over the control teaser focusing on the offer.

**G-: Free gift buckslip**
Removing the buckslip reduced response by 0.240 percentage points (6%). With this data, they could calculate a breakeven point for the full-color buckslip. For their best prospects the buckslip was profitable. Some less-responsive segments were more profitable without it.

**D+: Expiration date display (show through window)**
In addition to the second window (C+ above), a larger name/address window also helped. The offer expiration message, "Return this form within 10 days," was kept the same on the order form, but the envelope window was enlarged so the message showed through the window, above the recipient's name and address. This change increase response by 0.214 percentage points (5.35%).

**F+: Starburst on envelope back**
Adding the bright starburst with a "look inside…" message increased response by 0.194 percentage points (4.85%). In addition to the benefit of the starburst itself, the team realized that the back of the envelope can have a large impact on response and should not be ignored.

**H-: Lift note**
Eliminating the lift note was almost as harmful as removing the buckslip. Average response dropped 0.180 percentage points (4.5%) across the six mailings with no lift note. This made the copywriter happy and justified the minimal cost of the note.

## Step #5: Profit

**Implementing the optimal combination of all 11 elements, response rate jumped over 50%,** from 4.0% to 6.1%, and remained between 5.8% and 6.4% for subsequent mail drops. The test paid for itself the next drop when results were implemented.

In addition to the six significant elements, the team gained important insights from the five others. Changes to the letter (J and K) had no impact so the team kept their control letter, but realized that the changes they tested were, on hindsight, probably not very bold. The new order form, envelope colors, and background graphic also had no significant impact, so they kept these at the control.

### What If… they ran split-run tests instead?

Quantifying the effect of each element, the marketing team could keep the winning ideas and drop the bad ones, rapidly optimizing every element of the mailing. If these 11 elements had been tested as separate split-run tests, only one element—(C) the second window—would have been significant. They would have missed the five other significant effects (losing out on a 23% lift) and possibly implemented harmful changes like eliminating the buckslip and lift note.

Alternatively, they could have tested one element in each of the next 11 drops, taking one full year to implement final results.

### What If… they tested a few changes together?

It's also interesting to see how different their conclusions would have been if, instead of testing all elements in a scientific test, they had changed a number of variables in one package as a split-run test against the control.

For example, based on these results, if they had changed five things together—removing the buckslip (G+), showing the expiration date through the window (D+), adding the starburst (F+), changing the order form (I+), and testing short copy (J+)—they would have seen a 3.96% response rate for the test package versus the 4.0% response rate for the control.

They would have learned nothing or concluded that these changes should be avoided, even though two actually increase response (D+ and F+), two make no difference (I+ and J+), and only one is truly harmful (G+). At best, these results would have told them nothing useful. At worst, results would have led them to incorrect conclusions. Only by analyzing each element on its own could they gain valuable insights into what truly drives customer response.

## Step #6: Continue Testing – A Price and Offer Test

The consumer marketing team learned more in that one 11-element test than they had learned over the last two years of testing. Not only did they pinpoint how to increase response over 50%, but they also avoided making changes that had a negative impact on results.

The vice president wanted to keep testing. With price such an important element in every subscription offer, she wanted to focus on price-related variables in greater detail. They could do more creative tests in the future, but the next test included the five elements below. The consultant offered alternatives for testing more than two levels, but the vice president decided two bold levels for each element was best.

| | Test Elements | (–) Control | (+) New idea |
|---|---|---|---|
| A | Subscription price (annual) | $19.97 | $12.99 |
| B | Subscription period | 1 year | 18 months |
| C | Show cover price | No | Yes |
| D | Include 2 and 3-year subscription options | No | Yes, checkboxes |
| E | Expiration date format | Number of days | Specific date (6/19/03) |

**A: Subscription price (annual)** – The current subscription price was $19.97 for 12 monthly issues. Since increasing the number of subscribers was so important, they selected a lower $12.99 annual subscription price to test. The team hoped that the lower revenue/order would be offset by a large increase in the number of subscribers.

**B: Subscription period** – With one-year subscriptions, the marketing team sent out numerous renewal notices every year to keep customers coming back. They thought that a longer subscription period—18 months instead of 12—might not hurt response and could reduce marketing costs and increase the overall subscriber base over time. However, they wanted the per-issue price to remain the same whether offering a 12- or 18-month subscription. So for the test, the price (A) was adjusted depending on the subscription period (B):

- The current subscription price (A-) would be $19.97 for one year or $29.95 for 18 months
- The lower subscription price (A+) would be $12.99 for one year or $19.49 for 18 months

**C: Show cover price** – Sometimes they would show the annual cover price and sometimes not. They wanted to test whether it was useful information or got in the way of the offer.

**D: Include 2-year and 3-year subscription options** – Since subscription renewals required additional mailings and customer commitment, the marketing team wanted to add checkboxes offering customers the option to select a 2- or 3-year subscription at a lower per-issue price. If successful, multiple-year subscriptions would increase revenue (with the subscription paid up-front) and the long-term subscriber base.

**E: Expiration date format** – the last test showed that the expiration date increased response when it shows through the envelope window. In this next test, the team wanted to know if a specific date would be a stronger call-to-action, for example, "Return this form by 6/19/03," versus "… within 10 days."

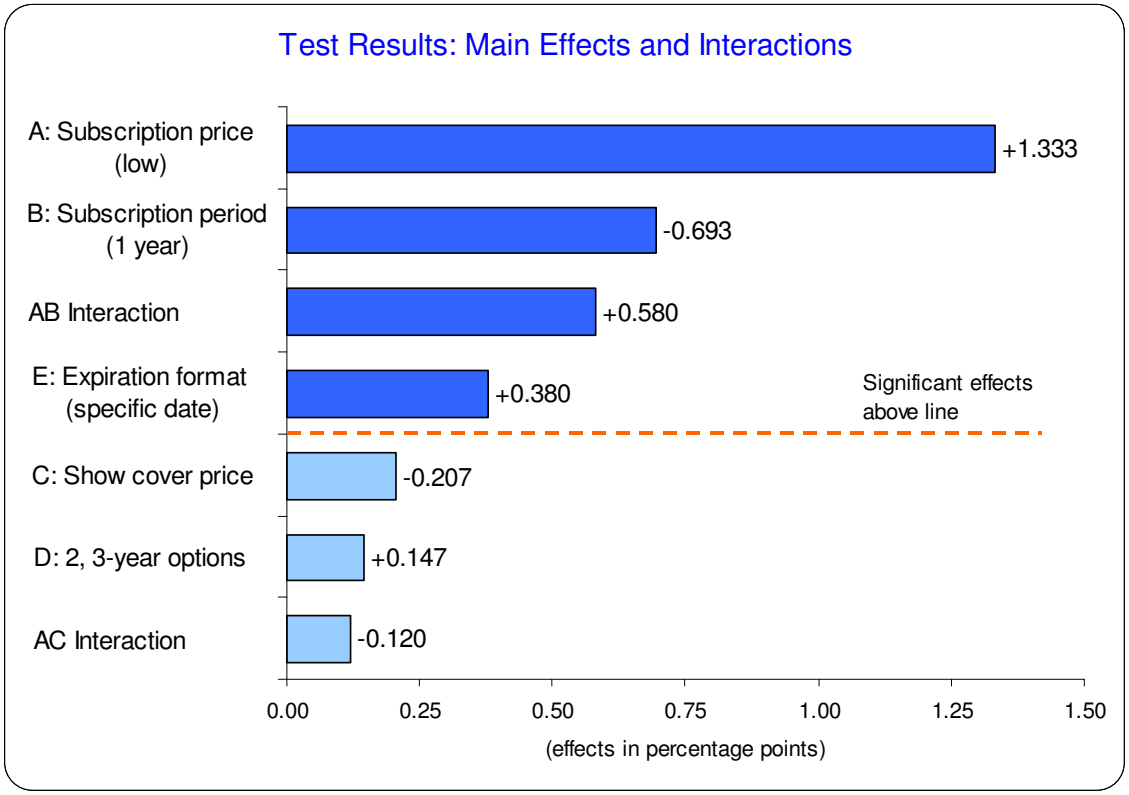
### Test Design

Focusing on five price and offer elements, interactions were more likely. The test expert suggested a different type of design that would more accurately identify and quantify potential interactions. Though larger test designs (i.e., more test recipes) permit the analysis of numerous interactions, the consultant suggested the 8-recipe test design, below. This test design offered the freedom to carefully analyze the most-likely interactions while limiting the number of test recipes.

| Recipes | A<br>Subscription price | B<br>Subscription period | C<br>Show cover price | D<br>2 and 3-year options | E<br>Expiration date format | Response |
|---|---|---|---|---|---|---|
| 1 | + | + | + | - | + | 6.61% |
| 2 | - | + | + | - | - | 4.44% |
| 3 | + | - | + | + | - | 6.49% |
| 4 | - | - | + | + | + | 6.24% |
| 5 | + | + | - | + | - | 6.71% |
| 6 | - | + | - | + | + | 5.05% |
| 7 | + | - | - | - | + | 7.05% |
| 8 | - | - | - | - | - | 5.80% |

Like the first 12-recipe test, this test design included the control (all minus) recipe as the last row. The other seven recipes were unique combinations of all five elements, providing the greatest amount of information in a small number of versions. With a smaller test—and with their experience from the last test—all eight recipes were created in just a few hours.

The team wanted to use a smaller sample size for this test, since price changes had some risk of hurting profitability. Analyzing alternatives, the group decided to use 60,000 names, split evenly among the eight recipes (7,500 names per recipe). In total, 3630 people responded, with response rates for each recipe listed in the matrix. Response rate for the control was 5.80%. Results are shown below.



## Test Results

With this test design, the five main effects and two interaction columns were independent. Three of the five main effects and one interaction were significant.

**A+: Subscription price (low)** – the $12.99 annual subscription price increased average response 1.33 percentage points over the higher $19.97 price (23% lift).

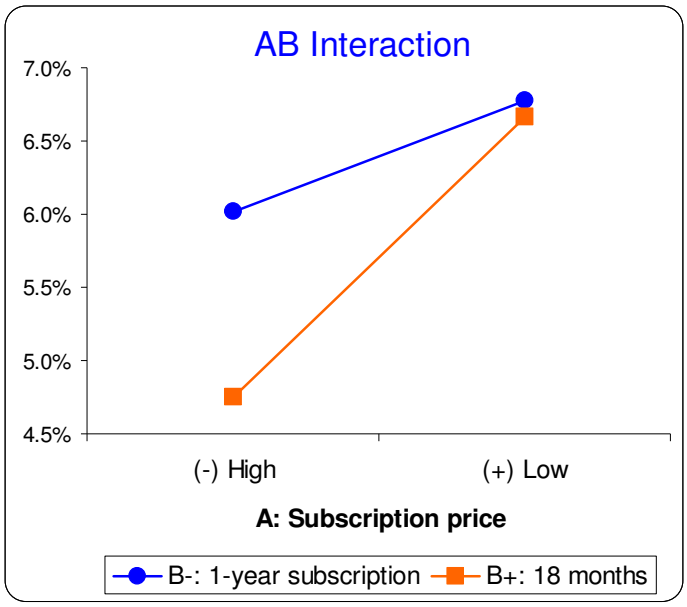
**B-: Subscription period (1 year)** – the longer, 18-month subscription reduced response, on average, 0.693 percentage points (12%).

**E+: Expiration date format (specific date)** – changing the wording of the offer expiration from "10 days" to a specific date increased response 0.380 percentage points (6.55%).

These effects are based on response rate, so additional profitability analyses (not shown) are also important. However, the most accurate data—and the most meaningful insights—came from looking at the interaction along with the main effects.

## AB Interaction

Interaction effects are often difficult to interpret… Is a positive effect good or bad? Since both A-B- and A+B+ give a positive AB interaction, which combination is best? As is often true, a simple picture can often elucidate the answers. The significant AB interaction is pictured below.




First of all, this AB interaction supports both main effects. The lower subscription price (both A+ points on the right) always has a higher response rate than the higher price (A- on the left). Also, the one-year subscription period (both B- points on the blue line) is better than the 18-month subscription (orange B+ line). However, an interaction occurs when these lines are not parallel.

This plot shows the very worst combination is A-B+: an 18-month subscription offer at the higher price ($29.95) is much worse than the main effects would predict. But look at the two points on the right…

When the subscription price is low, there is little difference in response rate between the 1-year and 18-month subscriptions. In fact, the low-priced 18-month subscription ($19.49) has a higher response rate than the high-priced 1-year subscription ($19.97).

In contrast to the results from both main effects, the AB interaction gives the marketing team a way to maintain per-subscription revenue while extending the subscription period to 18-months. This surprising result offers a new opportunity for growth, letting them:

- Increase response rate 18% by changing the expiration date format, lowering the per-issue subscription price, yet extending the subscription period to 18 months
- Maintain revenue per order
- In the long term, increase the subscriber base and reduce marketing costs by delaying renewals
- More accurately calculate profitability, combining data from the main effects and interaction

In addition, the non-significant effects were important. The 2- and 3-year options (D) did not hurt response and led to a few long-term subscribers, so the team decided the checkboxes could only help in the long run. Showing the cover price (C) had no impact, so the team decided to leave it off to avoid unnecessary clutter.

## Conclusions

These two scientific tests quantified the impact of 16 direct mail elements and led to a 70% increase in response rate. With growing profits and deeper insights, the team could see the marketplace more clearly and speed ahead in the right direction. Testing didn't solve all of their challenges, but it gave them the freedom to leverage their ideas, rapidly prove what sells, and stay on the leading edge of a highly competitive marketplace.