



Direct Mail Testing Unlearned: Myths, Methods, and What Actually Works





Abstract

I have been fascinated by the possibilities of allowing the market to tell you their preferences, not merely with focus-group opinions, psychometrics or surveys, but with decisions and dollars. I am also amazed at how few marketers are familiar with the highest levels of experimental design which direct marketing practitioners consistently achieve. Most importantly, while 'TESTING' has a wide array of meanings, the goal of 'predictive' marketing necessitates a more rigorous application of experimental design. This paper will focus on WHY direct marketers do not settle for influence testing but prefer aiming for hard financial predictions.

For more than 44 years I tried (with limited success) to distinguish valid predictive experimentation from 'give this a try' marketing. Suddenly, the validity of the casual attempts has become clear. Perhaps I was wrong to suggest that "throwing things against the wall to see what sticks" is not a valid example of testing. Accepting this as a valid use of the word illustrates the need to look deeper. This article attempts to summarize that work to clarify the meaning, objective and potential of this concept. The main differentiating factor is not in the testing itself, but in the goal and context of the tests.

Authored By
John Miglautsch

Contributions By
Morgan DiGiorgio

Sponsored By
DirectMail2.0



Introduction: Why Testing Matters

In the world of direct mail, one timeless truth stands above all: **"Always test something."**

Victor Hunter shared those words with me on my very first day in the industry over 40 years ago. It stuck. I've been preaching it ever since. But as I reflect on decades of experience, I've come to see that while many marketers **say** they test, very few do so with rigor. Worse, most don't know what scientifically valid testing even looks like.

Over the years, I've seen everything from "throw it at the wall" campaigns to carefully crafted experiments that rival pharmaceutical trials in their structure. And while the former can work, it rarely teaches you anything. That's the distinction I hope to make clear: testing can be more than just seeing what happens. It's about learning what works and **why**—so you can predictably improve your marketing efforts.

Let's clarify what testing is, what it isn't, and how direct mail testing—with or without digital channels—can become a reliable engine for business growth.





What is Testing—Really?

The word "testing" is used so casually that it has nearly lost its meaning. You can test bathwater, test a friend's patience, test a subject line, or even test a social media post just to "see what happens."

But when your CFO or CEO asks, "How do you know this will work again?" — that's where rigorous testing shines. That's when you realize that good marketing testing must go beyond anecdotal results. It must be **predictive**.

I learned this lesson in my early years working with major mailers. We didn't just test headlines or colors. We tested pricing strategies, catalog formats, and product sequencing. Sometimes, results changed our entire business model.

Too often, modern marketers confuse "engagement" with results. If a post gets likes, it must have worked, right? But engagement is not ROI. Financial executives don't want likes; they want lift, revenue, and profit.

So, let's strip "testing" back to its core. The essence of true testing is a cause-and-effect relationship.

If we do X, what happens to Y?

More importantly, will it happen again next time?



The Scientific Method of Marketing Testing

High-level testing isn't a guessing game. It's structured inquiry. That structure comes from the scientific method, something I became obsessed with after taking (and dropping) an advanced Organic Chemistry class back in college. One lecture on xenon fluorine reactions completely changed my outlook. The professor asked whether a scientist should publish a once-in-a-lifetime result if the experiment couldn't be repeated because it was knocked over by a janitor.

High-level testing isn't a guessing game. It's structured inquiry. That structure comes from the scientific method, something I became obsessed with after taking (and dropping) an advanced Organic Chemistry class back in college. One lecture on xenon fluorine reactions completely changed my outlook. The professor asked whether a scientist should publish a once-in-a-lifetime



Observation

You see something unusual: maybe a spike in sales from a weird ZIP code or unexpected success from a new list. Two of the best sources for new insights are customer feedback and data analytics and modeling.



Hypothesis

You form a testable idea:
"Would our list of gardening tool customers buy our cleaning products?"



Experiment

You split your audience randomly and send different versions.



Theory

You build a repeatable system based on tested truths.



Explanation

You interpret why one version outperformed the other.

I cannot stress this enough: your hypothesis must be specific, falsifiable, and grounded in prior observations. It must lead to clear yes/no answers, not vague assumptions.

Keys to a Valid A/B Direct Mail Test

I've seen more A/B test disasters than I care to count. One client mailed 150,000 pieces and got 3 responses. Statistically valid? Maybe. Useful Not a chance.

Another time, a client had 100 names and got 50 responses. That's a 50% response rate! Amazing, right? Nope. Too small to trust.

One of my favorite stories from the old days of cheshire labels: We'd use a room sized guillotine cutter to literally slice a mailing list printout in half—to randomize it. Then if we forgot to 'key' the mailing piece, we could take a juicy marker to the labels to track versions. Primitive? Yes. Effective? Absolutely.

While most of the foundational tracking techniques were perfected long before the Internet, as times and consumer buying methods have changed, it has become essential to adapt testing attribution.

In the 80's only 50% of consumers bought remotely at all. Many did not have a credit card. Most mailings included an order form or response card. Printing ink-jet technology allowed us to connect those cards with the individual addressees. With the advent of Wide Area Telephone Service (WATS) numbers, consumers were encouraged to call in their orders.

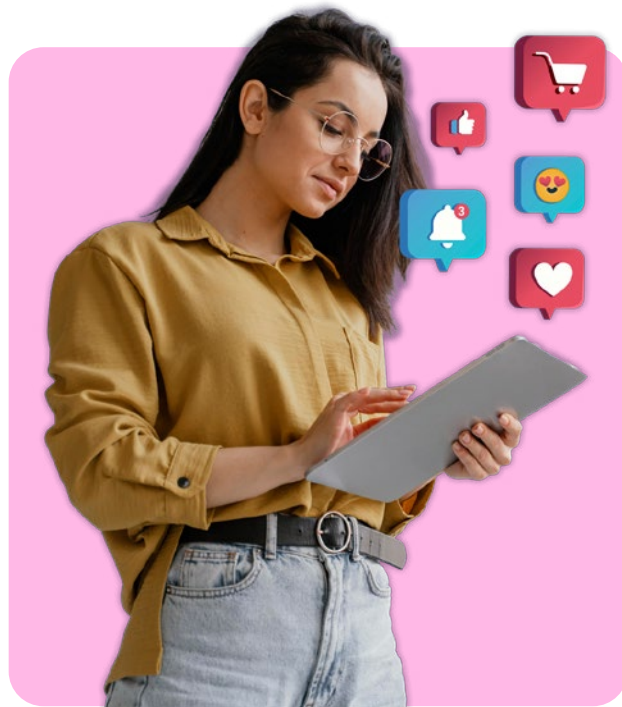
Comparison indicated that phone orders were significantly larger, perhaps due to the upsell opportunity the conversation afforded. This however broke some of the connection between the order and the response device. This disconnect has increased with the advent of the Internet. Virtually all response-generating offers encourage consumers to visit a website to respond.



Digital media makes 'clicking' to a link extremely convenient. Direct Mail physicality requires responders to key in, call, or more recently, scan a QR code. In this hand off, it is highly likely that the hard connection between the ad viewer and the tracking becomes disconnected. Web analytics makes significant assumptions about what prompted the response.

If, for example a recipient looks at a mailing piece and merely types or says the company name, google will offer a link. The consumer can continue to the offer and is happy. But from a tracking standpoint, it appears (especially to Google) as if the SEO or pay-per-click from the digital world prompted the action and is entirely credited to Google.

This issue is especially serious for companies who are early in the evaluation of Direct Mail in their marketing channels. There are many ways to adjust for this skew and stay closer to identifying the CAUSE of the consumer response, but it must be considered as part of your test planning process.



Clear Hypothesis

What exactly are you testing, and what do you expect to see?



Randomized Samples

Ensure no bias in how people are split.



Control Group

You must test against a known baseline, not two unknowns.



Sufficient Sample Size

You need enough responses to detect meaningful differences.



Tracking

Use codes, URLs, call centers—whatever it takes to tie response to source.

Direct Mail Testing In Action

CASE STUDIES





The Garden Weasel—When List Tells Product

Bon Ami was a legacy brand—a name your grandmother might’ve trusted. Known for its gentle powdered cleanser and that iconic yellow chick, Bon Ami had a small but loyal following. At the time, the company also owned the rights to the **Garden Weasel**—a successful “As Seen on TV” garden tool that had a cult following among hobbyist gardeners.

Bon Ami’s leadership attended a catalog conference and came back inspired. They wanted to modernize. Their plan? Launch a full product catalog featuring Bon Ami-branded cleaning supplies. And since they owned the Garden Weasel, they decided to mail the catalog to its buyer list.

As someone who’s spent decades helping brands avoid self-sabotage, I had to speak up.

(Actually in this case, I was called in after they had created their BonAmi catalog and been mailing it to the Garden weasel buyers. They were working with Jack Schmidt in KC. He gave me a call and said he thought I could help.)

I said, “Let’s pause. You’re sending a cleaning supply catalog to gardeners, expecting them to be excited because they bought a weeding tool once? You own a goldmine of customer data—you need to test what those customers actually want.”

(I actually asked if any other mailers were repeat users of the Garden Weasel list. They said several used as many names as they put up. But those repeat users were all garden related)

We launched a simple A/B test. Half the list received the cleaning-focused catalog as planned (the control). The other half (the test group) got a catalog curated around garden-related products—some new tools, accessories, and even outdoor décor.

The difference in response was immediate and unambiguous. The gardening-focused catalog significantly outperformed the cleanser one on both order response and average order value.

Here’s the deeper insight: the cleaning catalog wasn’t a bad idea. It was just a bad match for the audience. The mailing list—the Garden Weasel buyers—had already told us what they cared about. We simply weren’t listening until the test made it impossible to ignore.

This experience reinforced one of the most overlooked truths in direct marketing: It’s not about what you want to sell. It’s about what your audience is willing to buy.



The Takeaway: Testing revealed that Bon Ami had a more valuable asset in their audience than in their original product focus. By shifting their merchandising to reflect customer interest—and proving it through test-backed data—they not only saved the catalog but unlocked an entirely new product direction.

Back in the heyday of direct mail catalogs, Musician's Friend stood out as one of the most aggressive and consistent mailers in the music gear space. When we started with them, they were mailing over six million catalogs annually to musicians across the U.S. They sold everything from guitars to drumsticks, from cables to amps.

But there was a catch: Musician's Friend was owned by Guitar Center, a retail-first company that didn't love the idea of spending millions on paper & postage when they could drive traffic to their brick-and-mortar stores with less-expensive digital.

Every year, the catalog came under scrutiny from the top. And every year, we had to defend it—this time, with real data. When financial investors are involved, they care deeply about the rigor of the test design, the financial implications and its predictive force.

We proposed a clean hold-out test. From the upcoming catalog drop, we excluded 30,000 customers across all performance tiers: high, mid, and low value. These customers were randomly selected to not receive the catalog—while the other 1.5 million did.

We then watched and waited.

What happened was crystal clear:

The mailed group generated almost double the revenue per customer compared to the unmailed holdout.

The lift was consistent across value tiers: even lower-value customers showed noticeable increases in engagement and purchase frequency when mailed.

We didn't just prove that catalogs worked—we proved that not mailing would cost millions in lost sales.

And we didn't stop there. We replicated the test six years in a row, and every time, the catalog demonstrated similar lift. This wasn't a fluke—it was evidence. The catalog didn't just support sales; it sustained the business.



The Takeaway: In a climate where marketing teams must justify every dollar, Musician's Friend became a case study in how deliberate, long-term testing can silence skeptics and preserve valuable channels. It also highlights how direct mail, when executed properly, acts as a predictable revenue driver—even in a digitally noisy, multi-channel world.

Lovesac is known for its premium modular furniture and “Sactionals”—luxurious, reconfigurable couches that cost thousands of dollars. Early on, the brand invested heavily in physical retail showrooms, paired with a well-designed catalog program. But as digital channels became more dominant, executives began to question whether direct mail still had a place in their omnichannel strategy.

The CFO, looking at cost-per-catalog and the rise of digital media, had serious doubts. A new VP of Marketing came on board and asked a thoughtful, if familiar, question:

“Wouldn’t it be more efficient to stop mailing catalogs to existing customers—people who already know our name—and redirect those funds toward broader awareness?”

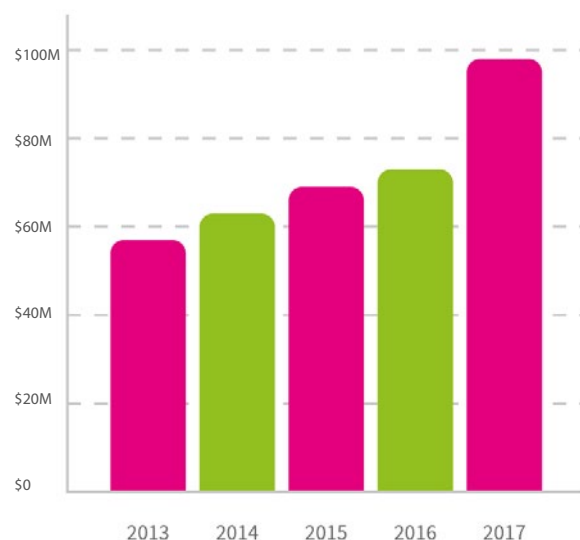
I responded with a question of my own: ***“Would you like to know the answer?”***

With that, we set up a classic hold-out test. We randomly selected 15,000 customers from their house list to not receive the catalog mailing. The remaining customers (the mailed group) received the full catalog drop as usual.

Over the next few months, we tracked all purchases—online, in-store, via phone, via QR code—and matched them back to the original customer lists. We weren’t just looking at attribution windows; we were measuring actual sales behavior by named individuals.

The results were astounding:

- The mailed group outperformed the unmailed group by \$30 per person in incremental sales.
- When we factored in cost of goods, overhead, and mailing expenses, the ROI on the catalog alone was over 300%.
- But here’s where it got interesting — neighboring households (those not directly on the mailing list, but often literally next door) also bought from the catalog.
- When we factored in this pass-along secondary halo effect—often neglected in attribution models—the ROI rose to 900%.



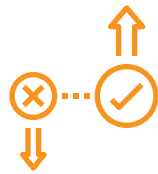
More astonishingly, for every 1 order from an existing customer, the campaign also generated 2 new customers—individuals not originally on the mail file but likely influenced by catalog sharing, which triggered showroom visits, or organic search following catalog discovery.

This wasn't just a win for direct mail. It was a masterclass in how mail can trigger a multichannel lift. What we proved—beyond a doubt—is that direct mail doesn't compete with digital or retail. It amplifies them.

Following this successful test, the direct mail channel was expanded to identify the incremental lift differential between 4 main benefits of their furniture. Digital post cards were used to drive traffic and also measure downstream EBITDA level ROI. Each customer drop was divided between 4 headlines. One benefit headline drove more profit than the other 3 combined. Once this finding was confirmed, that promotional message was rolled out in-store, digital ads and mass media. Same-store sales grew over 33% from the previous year.



The Takeaway: Lovesac's success wasn't about the catalog alone. It was about using direct mail as a **coordinated, trackable ignition point** for a larger omnichannel experience. With the right hold-out design and unified tracking across channels. Lovesac continues to mail existing customers as well as trade area households to this day.



Top 10 Omnichannel Testing Mistakes and How to Fix Them

Testing mail-only vs. mail + digital isn't hard, but it is easy to get wrong. The same few mistakes keep sinking otherwise good tests: uneven audiences, leaky controls, messy tracking, and conclusions based on tiny, rushed reads.

This quick guide calls out the top pitfalls I see most often and the exact fixes to keep your results clean, causal, and scalable so when you find lift, you know it's real and you can bet bigger with confidence.



1 **Unbalanced Audience Splits**

One of the biggest killers of a clean test is when the groups aren't truly comparable. If one cell skews heavily toward best customers and the other has more prospects, your results will reflect that imbalance.

To avoid this, segment your audience first with customers vs. prospects, active vs. lapsed, different geographies and then randomize within each segment. This ensures both sides of the test have a fair mix and any differences you see are caused by the variable, not the audience.

2 **No True Holdout**

A "control group" isn't really a control if it's still being hit by digital through retargeting, social ads, or brand search. When that happens, the lift from digital gets blurred across both groups, and you lose the ability to measure it.

The fix is to build a true mail-only holdout and hard-exclude those IDs from every platform and pixel container. That way, you know the control is truly digital-free and the results you measure are accurate.

3 **Too Many Variables at Once**

If you change the offer, the creative, the list, and add digital all in one test, you'll never know what caused the difference in performance. The results will be impossible to interpret.

Keep things clean by changing one major variable: whether digital ads are layered on or not. Everything else audience, offer, creative should stay consistent across cells. That's the only way to pinpoint the impact of the digital overlay.

4

Weak Tracking and Vague Destinations

Another common mistake is sending everyone to the homepage or using the same QR code, URL, or phone number for all responses. This makes it impossible to tie results back to the right test cell.

Instead, give each group its own unique tracking mechanisms: campaign-specific PURLs, QR codes, vanity URLs, or phone numbers. Even better, route the groups to distinct landing pages with clear tagging so you can see exactly who came from where.

5

Pixel Bleed on Shared Sites

If both cells hit the same website with the same global tracking tags, your control will get caught in the same retargeting nets as your treatment. That contaminates the data and destroys the integrity of the test.

To fix it, set up separate tag containers or workspaces, and suppress digital tags for the control group. Server-side gating by cell parameter is another clean way to prevent pixels from firing for those who shouldn't see them.



6

Tiny Samples and Short Windows

Testing on a handful of people for just a week gives you results that look directional at best, but they aren't statistically reliable. Small numbers can swing wildly and lead you to the wrong conclusion.

The solution is to size your test properly and big enough to detect meaningful differences and run it long enough to capture the full response window for mail. Only then will your results be solid enough to trust and replicate.

7

Timing Mismatches

If the digital campaign launches before the mail hits homes, or if one cell overlaps with a holiday or big promo period, your results are immediately compromised. The test ends up measuring timing differences instead of channel lift.

The fix is to lock calendars up front, making sure the in-home mail dates, digital flights, and blackout periods all align. That way, timing doesn't bias your results.

8

Platform Delivery Drift

Ad platforms are built to find easy wins. Left unchecked, they'll over-deliver digital impressions to the most responsive parts of your audience, skewing the treatment group's makeup mid-flight. That shift makes it hard to know whether lift came from digital or just a better audience mix.

Prevent this by using audience mutual-exclusion lists, setting frequency caps, and pacing controls. Keep an eye on delivery by segment to make sure the mix stays consistent throughout the test.

9

Last-Click Thinking

Looking only at the last click or last touch that closed the sale is a classic mistake. It ignores the fact that mail often acts as the ignition point that drives customers to the site, even if digital closes the loop. If you only credit last touch, mail's impact gets buried.

The better way is to attribute results by the randomized assignment itself reporting outcomes per assigned name or household. This keeps the test tied to causation, not correlation.

10

One-and-Done Conclusions

Finally, one of the most damaging mistakes is running a test once and assuming the results are universal. A single test is just a snapshot; it might reflect seasonality, audience quirks, or even random chance. Instead, re-run tests across different drops, seasons, and segments.

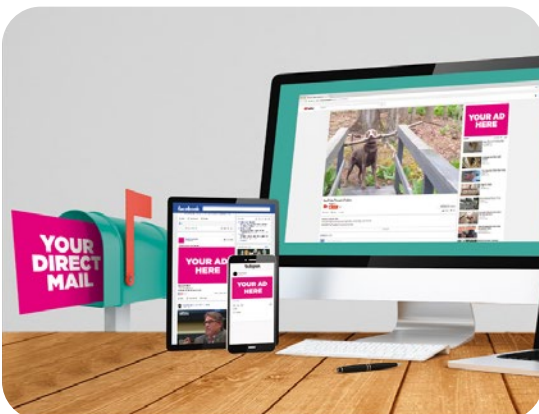
Look for patterns that hold consistently and note where digital adds the most value like with prospects, lapsed customers, or certain geographies. That's when you can move from "interesting" to "actionable" with confidence.





The big takeaway is that testing is only as strong as the design behind it. When audiences are balanced, controls are truly clean, tracking is airtight, and timing is aligned, you can trust what the data tells you. Skip those steps, and you're just guessing. But when you avoid these ten common mistakes, your tests stop being directional and start being decision ready giving you the clarity to prove lift, earn budget, and scale with confidence.

And here's one from the trenches: a digital marketer once told me, "If my ROAS is positive, I'm good." I asked, "Did you include product cost? Shipping? Overhead?" He stared like I had three heads.



Testing should build business cases, not just creative awards.

Perhaps the greatest pitfall, touched on earlier, is mistaken attribution. I regularly see digital marketers test a new channel, could be TV, Radio, OOH or Direct Mail. Each of those require responders to end up at the Web. Often, responders have visited before... checking out the products or company. They picked up a tracking cookie earlier and so that 'source' receives full credit. This is fine, except the direct mail channel, though actually triggering this visit and perhaps driving a sale, becomes invisible.

Another mistake which can be easily remedied, is using vague, unaccountable links in test. This can happen a number of ways, like directing people to a main home page rather than something offer specific, only having one landing page for various test offers. Even when hard links would be technologically simple, marketers do not direct QR codes through a tracking server and/or do not embed specific pixel tracking on key offer pages.



Testing in the Real World: Getting Started

You don't need millions of names. One test we did used odd vs. even ZIP code digits to assign test and control. No fancy software. No AI. Just logic and a spreadsheet. **Start with:**



Your Goal

Are you optimizing profit?
Testing new products?
Reducing churn?



Your Benchmark

Always have a baseline to
compare against.



One Variable

Change one thing at
a time: list, offer,
product, creative.



Long-Term View

Look for effects beyond the
first order. Thank-you inserts,
for example, increased repeat
purchases months later.



Final Thoughts

Testing is more than a tactic. It's a mindset. When done correctly, it transforms marketing from guesswork into a reliable growth engine.

We direct marketers have an edge. We can track. We can measure. We can know. But only if we treat testing as the serious discipline it is.

I hope this white paper encourages you to elevate your testing game. Because the truth is, when you test with purpose, your marketing becomes not just smarter—but unstoppable.



About the Author

John Miglautsch is a marketing futurist, philosopher, and founder of Miglautsch Marketing, Inc. With 40+ years in the direct mail trenches, he blends hard data with hard-won wisdom. He also holds an MA in Philosophy of Science and enjoys refuting 18th-century philosophers in his spare time.

Top 10 Omnichannel Testing Mistakes and How to Fix Them section was contributed by Morgan DeGiorgio, Chief Revenue Officer at DirectMail2.0 & Who's Mailing What!



Bibliography:

- Burnett, E. (n.d.). Ed Burnett's master guide to direct marketing. [Publisher not specified].
- Goldberg, B. A., & Emerick, T. (1990). Business-to-business direct marketing. Yardley, PA: Direct Marketing Publishers.
- Hunter, V. L., & Tietzen, D. (1997). Business to business marketing: Creating a community of customers. New York, NY: McGraw-Hill.
- Kobs, J. (1991). Profitable direct marketing (2nd ed.). McGraw-Hill.
- Miglautsch, J. (2012). Spinning straw into gold: An executive guide to the magic of turning data into money. (2nd ed). Seattle, WA. Kindle Publishing.
- Shaver, D. (1996). The next step in database marketing: Consumer guided marketing. New York, NY: Wiley.
- Stone, B., & Jacobs, R. (2001). Successful direct marketing methods (7th ed.). McGraw-Hill.
- Wunderman, L. (1996). Being direct: Making advertising pay. New York, NY: Random House.