Chapter 8.2

Testing, testing, testing

This chapter includes:

- ❑ What is testing?
- ❑ Why test?
- ❑ The first rule of testing
- ❑ The seven key marketing variables
- ❑ How to design a test programme
- ❑ The statistical principles of testing
- ❑ Samples and sampling methods
- ❑ Testing for smaller users
- ❑ Testing when response rates are low
- ❑ The cost of testing
- ❑ Monitoring and evaluating test results
- ❑ Press testing: a summary
- ❑ Testing in non-print media: DRTV, iTV, Radio
- ❑ Online digital media

About this chapter

In this chapter we look at testing, across *all* direct marketing media; its uses, benefits and methodology – and the science which underpins it. Each medium has its own technical complexities which we explore, but the fundamental principles are the same for all – from direct mail to email and SMS, from press to radio to DRTV, to telephone.

Today the old conflict between 'pure' research and 'real life' testing has all but evaporated. Modern direct marketers appreciate that the two disciplines are both complementary and indispensable: whereas market research seeks to tell us why people behave in certain ways, testing tells us how they actually behave and, more to the point, how they are likely to behave.

Robin Fairlie F IDM

This chapter has been written by Robin Fairlie and David Hughes, using also a quantity of material provided by Terry Forshaw.

Robin Fairlie graduated in history from Cambridge and joined the infant computer industry, leaving to become the first computer manager for Reader's Digest, and eventually serving on the Digest's UK Board. In 1980 he became Managing Director of the country's largest library supplier; in 1983 he was asked by the Post Office to serve as Chief Executive of the Direct Mail Services Standards Board, which aimed to improve ethical and professional standards in the direct mail industry.

In 1987 Robin became a freelance consultant. In this role he procured agreement between the Post Office and the industry on the detailed implementation of Mailsort, negotiated the formation of the DMA, and introduced the Mailing Standards Levy, which pays for the direct mail work of the ASA and for the Mailing Preference Service.

Robin helped to found the Direct Marketing Centre (predecessor of the IDM) and lectured on its diploma courses for many years. He has written two books and many articles on direct marketing, besides lecturing on MBA courses in the Far East. He is a Fellow of the IDM, the author of the Institute's Code of Conduct, and currently serves as co-editor in chief of the IDM's international Journal of Direct Data and Digital Marketing Practice.

Robin Fairlie, Tel 020-8340-2528;
e-mail: robin@idms.co.uk

David Hughes Dip DM, M IDM

David Hughes runs an independent digital marketing consultancy, Non-Line Marketing, and has over 8 years experience of email and internet marketing. His areas of specialisation include on-line data gathering, developing email marketing communication programmes and integrating email marketing with traditional direct marketing channels.

For the past 5 years he has been a Strategic Consultant with Emailvision, a European email technology provider, helping client and agency partners with the strategic, technical and operational issues affecting email campaign success. He works with companies such as Telegraph, Kangol, Institute of Directors and 3M and has experience of email in both the BtoB and BtoC sectors. David is also the Director of the Tank! On-Line Marketing Action Group, delivering thought leadership programmes for the client marketing community and is author of the DMA UK Email Marketing Benchmarking Report.

David spent 7 years at Claritas (now Acxiom) as the European Data Acquisition Manager generating over 8 million survey responses per year across 7 countries, both on and offline. He gained an IDM Diploma, with Distinction, in 1992, is a regular speaker for the Direct Marketing Association on email issues and is a member of the IDM Digital Marketing Examination Board and a Course Tutor for the IDM Digital Marketing Diploma.

davidhughes@nonlinemarketing.com
01737 218171

Chapter 8.2

Testing, testing, testing

What is testing?

Testing is the activity that sets direct marketing apart from all other forms of marketing. All marketing is about communicating with individuals; direct marketing alone make it the foremost aim to elicit a *response*, and offers the tools which enable the marketer to measure the cost-effectiveness of that response.

But we can go one stage further: using these tools, marketers can ask *questions* about product and marketplace (testing); provide *answers* to these questions (measurement) and make statistically reliable *predictions*. These three functions – testing, measuring and predicting – are the subject of this chapter, and they lie at the heart of any proper understanding of direct marketing.

Despite its unsophisticated label, testing is a sophisticated form of research. At its simplest, testing means a direct one-to-one comparison between two marketing options, e.g. between a high price and a low price.

For example, by presenting two different offers to identical samples of a target audience, an exact measure of the selling power of each offer can be obtained. Similarly, by presenting an identical offer to two different market segments, the responsiveness of each audience can be measured.

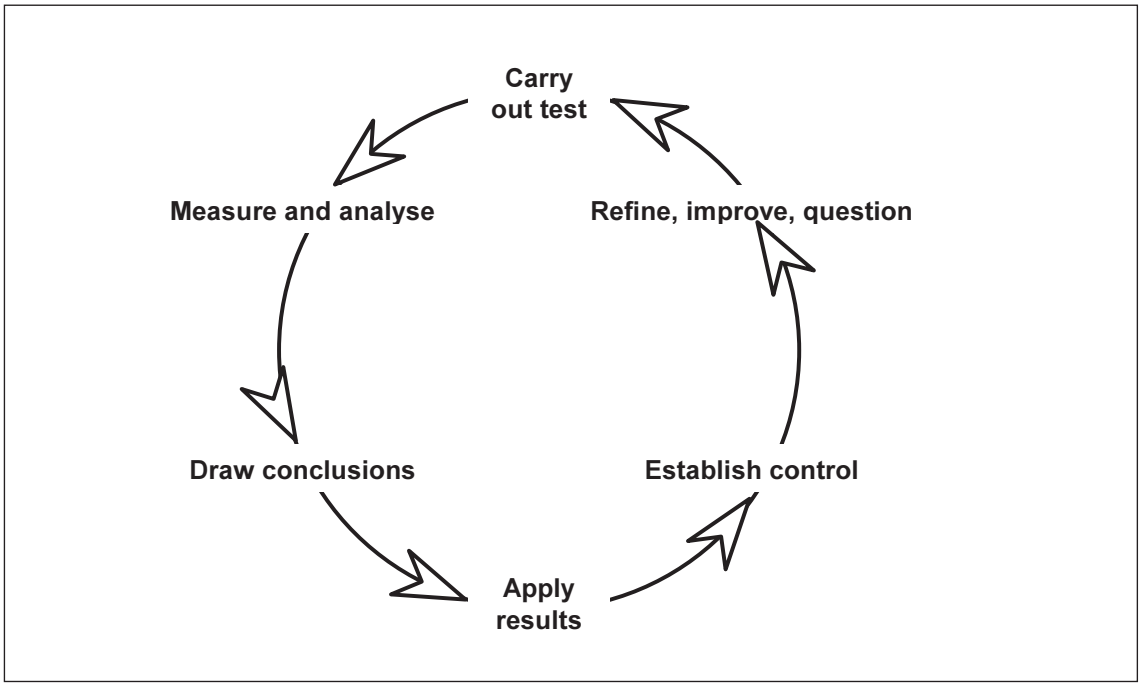
The simplest test of all would be to compare a single isolated variant with a known control. The example below in table 8.2.1 shows how a simple test programme can be developed to evaluate marketing options:

Table 8.2.1 **Simple test programme**

Control	List A	Proven offer
Test 1	List A	New offer
Test 2	List B	Proven offer (as control)

A simple set of tests such as that above, sheds light on a potential new market (list B) and a new offer (e.g. a lower price or a new incentive). So long as the samples are representative and the sample sizes are sufficient, the proper measurement of test results can predict the likely outcome of future marketing activity.
But testing is not a one-off event; it is a way of life, and should be conducted on a cyclic basis:

Figure 8.2.1 **Testing – a source of continuous improvement**



Why test?

If it ain't broke...

A memorable way of looking at the benefits of testing is to turn that popular adage "If it ain't broke, don't fix it" on its head. As Graeme McCorkell has pointed out, the huge advances made in Formula 1 racing car technology would never have occurred if things that hadn't broken were never replaced. Science advances precisely because innovators constantly pose the question: "How can we improve what is already working very nicely?"

So in direct marketing the rule is:

If it ain't broke yet, it will be one day. So study now how to fix it so it works better still and is even less likely to break in the future.

The first rule of testing

Many tyro direct marketers – and non-direct marketers with a smattering of knowledge – become preoccupied with their new toy: measurable response. All too often they channel all their creative energies into formulating (and testing) relatively unimportant variables, almost always those involving creative elements, e.g. copy style, layout and design, paper folds and envelope colours. Such tinkering rarely, if ever, pays off.

The golden rule of testing is that tests must be meaningful. *Only significant marketing variables should be tested*. As one of the modern industry's founding fathers puts it:

> "Test the BIG things"
>
> *Bob Stone,* Successful Direct Marketing

Testing *marketing* variables will normally have a much more dramatic impact on response and profitability than any esoteric creative test, and your test programme should reflect this.

The seven key marketing variables

The following seven factors in table 8.2.2 are the key elements responsible for the success or failure of *all* direct marketing ventures, regardless of medium. Later in this chapter we deal with specific issues relating to press advertising, to DRTV, to radio, to web advertising and to email. However, the principles on which we are now embarked (which have their broadest application in the context of direct mail and email) are applicable to all these, and to telephone testing as well (which has the added advantage over other media of person-to-person contact, allowing assessment of *reasons* for testees' behaviour.) We can rank these key elements in the following order of importance:

Table 8.2.2 **The seven key marketing variables**

1.	Product/service – nature and quality of	
1.	Target audience (list, media)	equal 1st
3.	Offer (e.g. price, incentives, terms)	
4.	Format	
5.	Creative	
6.	Timing	
7.	Response mechanism	

Fortunately, this hierarchical order of importance corresponds with the normal chronology of planning and implementation, i.e. in a normal sequence, product and audience are planned first and creative treatment later. So there is no excuse for overlooking the most important variables.

However, because these so-called **'**leverage factors**'** describe incremental gains, over time less incremental gain will come from testing the higher-ranked variables and more incremental gain will come from testing lower-ranked variables.

Testing is not only about maximising response

Bear in mind, as we discuss the impact of each of the seven chief marketing variables, that their impact will vary greatly from one situation to another. Remember also that we are not simply concerned with which will deliver the highest response.

As well as response, our test plans should take into account:

- ✔ Potential cost savings, effect on profits
- ✔ Outlay costs, absolute expenditure – affordable or not
- ✔ Competitive pressures (change may be unavoidable)
- ✔ Risks incurred or avoided (not necessarily financial)
- ✔ Long-term implications (you cannot always go back)
- ✔ Overall marketing objectives (not always response and immediate profit

Matching test objectives to marketing objectives

Company A's marketing objective was to recruit 10,000 new clients at a cost per new client of £10 each; its list/media test programme was structured to identify new lists or media to generate new clients within this threshold using a proven offer/creative approach.

Company B's marketing objective was to reduce mailing costs by 10 per cent while maintaining volume. The variables it selected for testing included cheaper formats and fewer enclosures per mailing.

We now look at each of the seven key marketing variables in turn:

Key variable: product/service

In many situations the product or service on offer will be predetermined. Test programmes will be built around the other six key variables.

However, the benefits of testing can be profound in the area of product design and development. If the product or service has insufficient appeal then all else will fail, and small-scale product tests will have limited the cost and risk at the launch stage. More significantly, testing can take place prior to expensive product development. This may involve *dry testing*.

Using dry testing, a number of product concepts and promotions can be tested simultaneously. Only successful products need go forward to more expensive product development stages.

What is dry testing?

Dry testing means developing a product concept and promotional package without having produced the final product. Promotions must make clear to prospects that the product will not immediately be available and should not request payment until the order has been fulfilled. To overcome negative reactions, dry-test products are often offered on special terms, e.g. a pre-launch discount. Subsequent production can be based upon responses to dry tests. Beware! Dry testing can cause serious hostility among prospects if it is not fairly and openly conducted.

Key variable: target audience

The selection of lists and choice of media (and programme content in the case of TV and radio) are absolutely critical to the success of direct marketing. The wrong creative approach to the right audience may succeed, but the right creative approach to the wrong audience is doomed to failure.

Mailing list test options include:

- ✔ Customer lists
- ✔ Enquirers
- ✔ Lapsed or dormant customers
- ✔ Affinity groups
- ✔ Lifestyle databases
- ✔ Cold lists (i.e. bought or rented lists)

Other addressable media that can also be tested include:

- ✔ Phone
- ✔ Electronic media (e.g. email lists, visitors to websites, responders to SMS messages)

Key variable: offer and price

The best offer you'll ever make your customers is the intrinsic promise of your brand. But few brands are so strong that an additional offer won't greatly enhance a product's appeal and increase its sales. As Jim Kobs says, "*The right offer can sell almost anything*". The key word is 'right'. The offer needs to be relevant to the target market.

Offers may be as simple as the product at an attractive price. Or they may be multi-faceted including optional extras, special packaging, incentives, terms, payment options and guarantees etc.

Importantly, offers not only affect response, they also impact other variables, i.e. different offers produce different levels of ongoing customer performance. For this reason, sophisticated marketers track the results of offer tests for several years.

Price is often as critical as the product itself in terms of whether or not a sale is made. Its effect is something which most experienced exponents find impossible to predict without testing. For example, a high price can occasionally pull the best response, while a low price may not guarantee you a sufficient number of orders. For this reason, and because of the impact of price on profitability, price testing should be considered wherever possible. To non-marketers, the concept of price testing is often suspect, but if the exercise is carried out sensitively there should be no problem. After all, the aim of price testing is not to find the highest price at which goods can be sold; it is to find the *profit maximising* price and this may well be a lower price than customers might otherwise expect to pay.

In a direct high/low price test, it is important that customers buying the product at the higher price are refunded the difference and the reason for the lower price

should be explained, e.g. lower production costs resulting from higher volume responses.

If you test only two prices against each other, and the low price proves to be more profitable, what have you learnt? Perhaps a still lower price would do even better….So, always test three prices, with the one you consider most likely in the middle.

Key variable: format

The physical format of a mailing piece or advertisement is another important area for testing. For example, it has been known for a large C4 pack to more than double response when tested against a smaller DL pack. Conversely, small packs have triumphed over large packs in different situations, especially when their lower costs are taken into account.

Possible format tests include size, number of components, paper weights/ qualities, types of envelope (e.g. paper or polythene), and addressing method. The equivalent in email marketing might be length of message, text versus HTML, and so on.

Experience shows that the following tests are not usually significant: colour of paper, label versus direct addressing, blue versus black signature, name and style of signature, franking versus preprinted postal impression etc. – yet all of these are regularly tested by blinkered marketers at the expense of more meaningful marketing variables such as, for example, payment methods.

Key variable: creative treatment

Although low on the hierarchy list, creative elements (copy, layout, typography and colour etc.) can significantly affect results. However, care must be taken when drawing conclusions. A well-written letter clearly stating benefits, proposition, product details and action required may increase response, but it is easier to conclude *that* it works rather than *why* it works.

Whose garden is it anyway?

A manufacturer of garden tools decided to find out whether results were affected by the use of male or female gardeners in illustrations. To his surprise, although his garden implements were heavy and difficult to handle even for professional gardeners, both enquiries and sales were enhanced by some 20 per cent when female models were employed. This does not necessarily mean more gardeners are female, although it might mean that. It is a typical example of how testing shows what happens, while research is needed to know why it happens.

So what to test?

The list below demonstrates only a fraction of the number and variety of tests regularly carried out by direct marketers:

- ✔ Product, product name, with and without accessories, deluxe or standard versions
- ✔ High price versus low price, payment terms, paid or free trial, size of discount
- ✔ Free gift versus no free gift, nature and number of gifts, reason for gift
- ✔ Prize draw: nature and value of prize(s)
- ✔ Media tests, e.g. newspapers versus magazines, positions, sizes, timing
- ✔ Lists
- ✔ Formats, e.g. envelope size, number of components, use of colour
- ✔ Copy, tone, positioning, length

Key variable: timing

Although timing tests are often neglected in test programmes, timing can play a very significant role in direct marketing success and failure. There are obvious situations where the product has a link with seasonality, e.g. a gardening product, a Christmas hamper, a holiday or an educational product coinciding with the commencement of the academic year.

However, seasonality is often far more subtle. For example, when would you expect private motor insurance applications to peak – at the beginning of the registration year, in January, at the beginning of summer or none of these? Certainly until recently the peak was in March, one explanation being that motorists used to lay up their cars for the duration of the winter!

Clearly it is difficult to generalise about timing. Some mail order companies' tests have shown July to be a strong selling period – maybe as a result of less 'clutter' in the press and mail from competitors. Others typically favour January to February and September to November. Timing should therefore always be included in any serious test programme if results are to be maximised. Not surprisingly, media owners have become aware of these peaks and troughs and now adjust their rates accordingly, so that planning should take into account savings from using less popular, cheaper time slots.

One of the surest guides to seasonality and timing is directories such as Yellow Pages, where an advertisement may sit unchanged year in, year out. A simple tracking of enquiries/sales will accurately pinpoint peaks and troughs of interest.

Some like it hot

When would you expect swimming pool enquiries to peak? Remember a pool may take several months to plan and build and so, to be ready for the first hot day of summer, you might expect prospects to make their plans just after Christmas, or perhaps March at the latest. When do they peak? Many suppliers report enquiries peaking on the first hot day of summer! No doubt the same goes for other summer-related capital goods: sun awnings, conservatories and garden furniture etc.

Key variable: response mechanism

Time was when direct mail only offered coupon response. Tests of coupon therefore focused on the size of the coupon or creative considerations, e.g. whether or not to repeat the offer or call to action in the coupon.

Nowadays, however, there are many more ways your prospect might respond, notably by phone, fax or email. So, you might wish to test Freephone versus STD or fax *plus* email against fax *or* email.

In testing any of these combinations, though, remember the general rule of response: the more response options you offer, the higher response volumes are likely to be.

One final thought: important though response volumes and rates are, it is the cost of the response and its quality that will matter most. The volume of response is quickly measured, but the quality of the response takes much longer to determine.

How to design a test programme

Before we begin to design a test programme we need to fully understand the campaign of which it is to form a part.

In table 8.2.3 below a 12-step checklist for planning and executing a test programme has been compiled, from initial campaign objectives to final analysis and reporting on test results. We now explore the three practical steps in setting up a test programme: steps 4, 5 and 6 from that master checklist.

First establish the control

In direct marketing generally, the control is simply your best-performing package, insert, advertisement or script. It is the base against which all test results should be measured. The control is the approach you would use if you were unable to test alternatives; what some exponents term 'the banker'. Test programmes should monitor the continued effectiveness of the control while at the same time trying to beat it – the 'beat the champ' principle.

If yours is a new venture you will have to research industry norms to establish a control; using direct mail you will probably start with a fairly standard pack (e.g. letter, brochure, order form, outer and reply envelopes). This need not inhibit you from being more adventurous – it simply means that you will have a baseline against which to judge more ambitious departures. Hopefully you will conserve your creative energies for testing the key marketing variables!

An important reason for testing – perhaps *the* most important reason – is that at some point your control will tire and cease to perform adequately. At this point you will need a replacement control, developed and honed from your test programme.

Table 8.2.3 **The 12 steps in building a test programme**

Step 1	Define the overall campaign objectives.	What is important to the campaign's success? What are the *real* objectives?
Step 2	Understand the parameters of the campaign as a whole.	How many mailings? Available list size? Opportunities for testing? Budget? How responses to be handled? Time span? Is response device logical and clear? Data to be captured etc?
Step 3	Decide list selection criteria.	Remember test lists must be representative of the market universe. See Part III of this section.
Step 4	Establish the control/s.	See details in the text.
Step 5	Decide the test strategy.	See details in the text.
Step 6	Construct test matrix.	See details in the text.
Step 7	Carry out tests.	Is any cell responding far better than others? Can this be explained? Are all cells contacted at the same time? Is bad weather affecting overall response? Note your conclusions for future reference.
Step 8	Observe responses.	Was the programme executed as planned? If changed, will this affect results?
Step 9	Check back.	Plot key variables and record what worked and what failed to meet objectives.
Step 10	Analyse results.	Can you spot any correlations between winners and losers? Any key variables among responders to tests, e.g. age, gender?
Step 11	Study analysis.	Try to gain a subjective view of who is responding to tests, and why.
Step 12	If possible, undertake other forms of analysis, e.g. regression analysis.	Begin the process again.
Step 1	Define overall objectives for the next campaign.	

Next decide the test strategy:

Test strategy statement

Project name: SPRING 2001 WIDGET MAILING

Issued: 11/09/2000

Control: C4 control pack with 2-page letter, control brochure and reply card

	Test	Objectives	Method
1.	Format test: C5 versus C4 outer	To reduce mailing pack costs while maintaining control response	All internal elements as control pack, folded to fit C5 outer envelope
2.	Free gift added – lowkey presentation	To improve response and profitability by offering a free gift, without affecting the basic tone and structure of the control pack	Mention free gift in the letter copy and PS
3.	Free gift added – heavy emphasis	To improve response and profitability by offering a free gift and featuring the gift heavily throughout the pack	Include separate fourcolour 'free gift' flyer. Refer to gift in the letter and copy and PS and include an order card with illustration
4.	New creative	To achieve a 'breakthrough' against the existing control by using a very different approach	All new elements

An important rule when constructing a test strategy is: *do not test more than one variable against control at a time*. If, for instance you change, say, the offer and the creative in one test and you get a significant uplift in response, to what can you attribute that uplift, the offer or the creative?

There are, however, two major exceptions to the rule of single-variable testing:

1. In some situations it may be wise to break right away from earlier thinking and test a completely new approach, in which case the more different the new approach, the better. This should never be done in the absence of a strict control, or the effect in future years may be literally to lose one's bearings and not know which way to turn next ('wandering').

2. Multi-variable tests can be carried out if a strictly scientific basis is used to determine sample size and test structure, as we explore below.

Finally, construct your test matrix – direct mail and email only

At the most advanced level of direct marketing, where large amounts of money are at stake, multi-variable test matrices are constructed and monitored by statisticians. For the lay reader, some guidance is given below on compiling direct mail test matrices to compare several variables at once.

Let us suppose you want to compare three lists (A, B, C) and three offers (1, 2, 3). Strictly speaking you would need to cover all possible combinations of list and offer, each to a statistically significant sample. If your minimum sample size was 6,000, your test matrix would look like that in table 8.2.4 below. It would

consume 54,000 valuable names (always assuming that you had 54,000 names available!).

Table 8.2.4 **Test Matrix I**

	Offer 1	Offer 2	Offer 3	Total
List A	6,000	6,000	6,000	18,000
List B	6,000	6,000	6,000	18,000
List C	6,000	6,000	6,000	18,000
Total	18,000	18,000	18,000	54,000

Consuming too many names in a test, even if available, can be a serious error, and so an alternative matrix may be needed. It is possible to reduce the number of test names used by making some careful assumptions.

If we assume that the three lists (A, B, C) are sufficiently similar and the best offer is likely to work across all three lists (i.e. the lists and the offers are independent variables), then the test matrix can be revised as in table 8.2.5. Test Matrix II uses only five test samples and reduces the name requirement to 30,000 overall. The likely outcome of the blank cells can be extrapolated from those mailed.

Table 8.2.5 **Test Matrix II**

	Offer 1	Offer 2	Offer 3	Total
List A	6,000	6,000	6,000	18,000
List B	6,000	-	-	6,000
List C	6,000	-	-	6,000
Total	18,000	6,000	6,000	30,000

However, an even more efficient method is a *block design* which can be used to combine all the test variables in one matrix and further reduce the number of test names used. In Test Matrix III in table 8.2.6 below, only 18,000 names have been consumed. The minimum mailing quantities of 6,000 per test are arrived at by adding up the columns. The required results are the totals for each column, both horizontally and vertically.

Table 8.2.6 **Test Matrix III**

	Offer 1	Offer 2	Offer 3	Total
List A	2,000	2,000	2,000	6,000
List B	2,000	2,000	2,000	6,000
List C	2,000	2,000	2,000	6,000
Total	6,000	6,000	6,000	18,000

The benefits of multi-factor or block designs like the one above (Test Matrix III) are two-fold:

1. The total sample size is kept to a minimum.

2. Some light may be thrown on the effects of different combinations of factors. Remember, for these interactive effects to be statistically significant, each

individual cell total would need to be 6,000 names, which brings us back to Test Matrix I and a total of 54,000 names.

By keeping the numbers within each test cell to a minimum you run the risk of an occasional freak result. If, in a multi-factor test matrix comprising small volumes, the result of one of the cells is abnormally different from the other cells, it may be wise to ignore the result for that cell. The discrepancy may be caused by some operational problem and not the test variable itself – known as 'statistical error'. Alternatively you would need to investigate the 'freak' result very carefully and ascertain the likely cause.

Table 8.2.7 **Test Matrix IV**

	List 1	**List 2**	**List 3**
Offer 1	Copy A	Copy B	Copy C
Offer 2	Copy B	Copy C	Copy A
Offer 3	Copy C	Copy A	Copy B

For Test Matrix IV – a 'Latin square' – we see that *nine* variables have been included: three lists, three offers and three copy variants. To test these nine variables in straight 'head-to-head' tests, i.e. testing one element against each other at a time, for the results to be meaningful, would involve 36 paired comparisons and 216,000 names.

But by using a Latin square, these tests can be carried out with only 45,000 names, or 5,000 per cell, saving hundreds of thousands of pounds on the programme.

Note each offer, list and copy variant appears in three cells. To design and evaluate the results of a complex matrix like this requires either a software package or a statistician.

The statistical principles of testing

All prediction based on the outcome of tests relies on the theory of probability. The theory is used to assess the chances of a given test result being repeated.

By using statistical probability direct marketers can interpret response data and make decisions on the basis of objective criteria. Statistical principles provide a better decision-making foundation than 'gut feel' or partial experience. *By balancing statistical probability with sound judgement direct marketers can make the best possible decisions.*

The role of statistics

Statistics are used to *generalise*, e.g. to summarise the behaviour of customers in a particular target market. In direct marketing, the aim is normally to *measure* the behaviour of past and present customers in order to *predict* the behaviour of future customers, although the direct marketer should never forget that predictions based on test results are conditional on other factors remaining equal.

The three statistical concepts discussed in this chapter are:

1. Inference

2. Significance

3. Confidence

1. Inference: accuracy depends on sample size

Direct marketers not only generalise from past and present customers to future customers; they also use part of their target market to stand for their entire target market. This process of generalisation is called 'statistical inference'. The term 'inference' is used for good reason: it is never possible to prove that a future response rate will be X per cent. This can only be inferred. The smaller the group that is studied, the less certain one can be that the inference is a good one.

Because analysing the whole of a target market is usually expensive, time-consuming and often impractical, tests are performed on *samples*. A sample should be of sufficient quantity, and of a suitable character, to render it unbiased and wholly representative of a total 'universe'.

2. Significance: is that test difference meaningful?

Significance is what you are hoping for when you plan and execute a test: you want to know if one course of action is more productive than another, by how much, and – this part is often overlooked – whether the result is *statistically* significant. Not whether it is significant to you but whether it is significant scientifically, i.e. not simply the effect of chance. In other words, a statistically significant result is one that cannot be explained by sampling error.

3. Confidence: how reliable is the test overall?

Because all tests are probabilistic, we need one more assessment of the accuracy of a test before we can be sure of its value. This is the confidence level, which represents the degree of assurance that the test sample and its result are representative of the total market.

The probability that a test response truly lies within a given range is its *confidence level.* Most often 95 per cent confidence is the chosen level of confidence – implying that only once in 20 times will an identical campaign produce results outside the range inferred from your test. (There is of course, no such thing as a 100 per cent confidence level.)

As the marketer you must set the level of confidence required – that is, how confident you need to be of the significance of your test results.

Samples and sampling methods

Ideally, the job of selecting a proper test sample should be given to a trained statistician. For small direct mail practitioners, or users of email, that may not be feasible. They then, above all others, must understand clearly the statistical methodology on which good sampling depends.

Users who can rely on good professional statistical help should also, for their own protection, know the principles that underlie the method. Since this chapter is in

no sense a handbook of statistics, it is to these principles, together with some simplified rules of thumb, that we will now address ourselves.

The normal method of obtaining a random, representative sample in direct marketing – e.g. from a list of names and addresses – is to select every 'nth' name in the list. (You can safely ignore any jumped-up mathematician who tries to tell you this is not satisfactory.) Thus, if you have a list of 100,000 names, and you require a sample of 5,000, then you will take every 20th name for your sample.

How do we determine the required sample size? First, let's dispose of a common fallacy: it is sometimes alleged that sample size should be related to the size of the universe from which it is taken – 10 per cent or some such. This is nonsense. (If it were true, Gallup, in order to produce a valid sample of electors for a political opinion poll, would have to interview several million people.)

The criteria for determining minimum sample size are:

The confidence level required (usually 95 per cent)

The percentage variance above or below your observed test result which you will tolerate

The approximate expected response rate

That is, if the response rate is *c*% we wish to be *a*% certain that *c*% of our universe, plus or minus *b* will respond in the same way.

There is a completely generalised formula for working out the correct sample size (S) for all possible values of *a, b,* and *c* – but it is complex. So long as we are working with a 95 per cent confidence level, a simple formula will do:

$$S = \frac{3.8416 \times c\,(100 - c)}{b^2}$$

Now try an example. Imagine a list of prospects from whom we must get a response rate of 4.5% in order to break even; the best estimated response is 5% and the limit of error we can tolerate is 0.4%. Then our sample size must be:

$$S = \frac{3.8416 \times 5(100 - 5)}{0.4^2} = 11{,}405$$

But if we expect a better response – say 5.5% – we can afford a wider margin of error without risk of falling below break-even. Our formula might then become:

$$S = \frac{3.8416 \times 5.5(100 - 5.5)}{0.9^2} = 2{,}465$$

Always choose a value for *c* which errs on the high side, and a low value for *b*. This may give you a larger sample than necessary, but this may be better than ending with a statistically invalid result.

Once the *actual* response to your test is known, the formula can be transposed in order to indicate what the real margin of error in your result is. Thus:

$$b = \sqrt{\frac{3.8416 \times c(100 - c)}{S}}$$

Then, if the actual test response were 5.2%, with a sample of 11,405:

$$b = \sqrt{\frac{3.8416 \times 5.2(100 - 5.2)}{11{,}405}} = 0.41$$

I.e. we can be 95% certain that a mailing to the universe of which this sample is representative will (other things being equal) produce a response of 5.2% plus or minus 0.41 – that is, between 4.79 and 5.61%.

Or, with our smaller sample

$$b = \sqrt{\frac{3.8416 \times 5.2(100 - 5.2)}{2{,}465}} = 0.88, - \text{ a result between 4.32 and 6.08\%}$$

It is worth noting that, for most practical purposes we are really operating at a level of 97.5 per cent confidence rather than 95 per cent. This is because our variance can be either above or below our test result. Most practitioners worry about checking the low point of their expectations, which they can be 97.5 per cent sure of exceeding, and are less worried about the 2.5 per cent chance of getting a result above the maximum variance.

The foregoing examples show how to set sample sizes and project rollout responses from test samples. In practice, rollout conditions are rarely consistent with test conditions, and many factors need to be considered when assessing the viability of rolling out a test. For example, the timing of the rollout, the level of competitive activity, and general economic circumstances, could all significantly affect response. So be careful: this is where judgement must come to the aid of statistics!

Testing alternative approaches

So how do we determine statistically valid differences between alternative approaches, e.g. when more than one sample is being mailed?

Is the difference significant?

A test sample of 10,000 was mailed with Pack A and a further 10,000 received a variation that we shall call Pack B. The question is: are the different results statistically significant?

The response percentages were:	Pack A	2.14%	(+/- 0.283 – i.e. 1.857 – 2.423)
	Pack B	2.68%	(+/- 0.316 – i.e. 2.364 – 2.996)

Since there is potential overlap between these results (Pack A could go as high as 2.423%, while Pack B could be as low as 2.364), we cannot be 95% confident that Pack B will perform better. (We could, however, be 90% confident: rework the formula using 2.706 in place of 3.8416 for a 90% confidence level; the overlap now disappears.)

Using statistical tables

We have made much of the statistical formulae in this section to demonstrate the principles involved. In day-to-day direct marketing it is far more likely that you will refer to published tables of sample size and probability.

Below in table 8.2.8 you will find a useful table of sample sizes covering responses from 0.5 per cent to 10 per cent, the range most frequently found in consumer direct marketing. The table shows the minimum sample size required for different response rates and accuracy levels.

To use the statistical table, given the most commonly used 95 per cent confidence level, you need to have a broad idea of the response rates that you anticipate. This will probably be based on previous controls. As the minimum size of sample increases with higher response rates you should slightly overestimate your anticipated response levels when deciding on sample size.

Then, you must decide on an acceptable level of accuracy. The accuracy will affect any calculations of break-even or profitability. For example, are you going to be satisfied by results which are within + 0.5 per cent of your expected response, or do you need to be sure to within 0.2 per cent? The decision often depends on the lower limit as this may represent break-even or budget cut-off points.

Table 8.2.8 **Table of minimum sample size for a test or control mailing 95 per cent confidence level**

Anticipated response rate to the mailing	Acceptable plus or minus error on anticipated % response												
	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	1.0
0.5%	19,100	4,800	3,100	2,100	..	..	..	..	..	..	..	..
1.0%	38,000	9,500	6,100	4,200	2,400	..	..	..	..	..	..	..
1.5%	56,800	14,200	9,100	6,300	3,500	2,300	..	..	..	..	..	..
2.0%	75,300	18,800	12,000	8,400	4,700	3,000	2,100	..	..	..	..	..
2.5%	93,600	23,400	15,000	10,400	5,900	3,700	2,600	..	..	..	..	..
3.0%	111,800	27,900	17,900	12,400	7,000	4,500	3,100	2,300	2,000	..	..	..
3.5%	129,800	32,400	20,800	14,400	8,100	5,200	3,600	2,600	2,300	2,000	..	..
4.0%	147,500	36,900	23,600	16,400	9,200	5,900	4,100	3,000	2,600	2,300	..	..
4.5%	165,100	41,300	26,400	18,300	10,300	6,600	4,600	3,400	2,900	2,600	2,000	..
5.0%	182,500	45,600	29,200	20,300	11,400	7,300	5,100	3,700	3,200	2,900	2,300	..
5.5%	199,700	49,900	31,900	22,200	12,500	8,000	5,500	4,100	3,500	3,100	2,500	2,000
6.0%	216,700	54,200	34,700	24,100	13,500	8,700	6,000	4,400	3,900	3,400	2,700	2,200
6.5%	233,500	58,400	37,400	25,900	14,600	9,300	6,500	4,800	4,200	3,600	2,900	2,300
7.0%	250,100	62,500	40,000	27,800	15,600	10,000	6,900	5,100	4,400	3,900	3,100	2,500
7.5%	266,500	66,600	42,600	29,600	16,700	10,700	7,400	5,400	4,700	4,200	3,300	2,700
8.0%	282,700	70,700	45,200	31,400	17,700	11,300	7,900	5,800	5,000	4,400	3,500	2,800
8.5%	298,800	74,700	47,800	33,200	18,700	12,000	8,300	6,100	5,300	4,700	3,700	3,000
9.0%	314,600	78,700	50,300	35,000	19,700	12,600	8,700	6,400	5,600	4,900	3,900	3,100
9.5%	330,300	82,600	52,800	36,700	20,600	13,200	9,200	6,700	5,900	5,200	4,100	3,300
10.0%	345,700	86,400	55,300	38,400	21,600	13,800	9,600	7,100	6,100	5,400	4,300	3,500

Testing for smaller users

Many smaller users will, of course, never achieve the minimum sample sizes required, to be 90 or 95 per cent confident about their test results.

This does not mean that small operators should not test. On the contrary, testing is as important for small users of direct marketing as it is for the giants of the industry. It simply means that you should not accord too much credence to isolated tests. You should also not attempt to test minor departures from your control, but only major marketing differentials.

The golden rules for smaller operators include:

- ✔ Repeat important tests several times
- ✔ Validate test results by other research methods, e.g. follow-up interviews, customer feedback and staff reports
- ✔ Keep tests to absolutely significant differentials
- ✔ Test only one or two major departures per campaign/season
- ✔ Keep tests to criteria which cannot be verified by other means

Testing when response rates are low

There are, unfortunately, circumstances in which the mathematical formulae discussed above do not hold. This is the case when response rates are very low – below 0.1 per cent. This will not often occur in direct mail or email (although it might in the case of a very high priced product) but will be the case in most space advertising. (The same is also true in radio or TV advertising when you are testing the response from one region against the response from another – i.e. the audiences being tested are not matched samples.)

In these circumstances, the procedure for comparing the results of two tests (or a test and control) is as follows:

1. Total the responses from the two tests
2. Determine the confidence level (90 or 95 per cent) required
3. Then, for the difference between the two responses to be significant, the larger response rate must constitute a proportion of the total as shown in the table below:

Table 8.2.9 **Significant percentage share for confidence level of:**

Total response	90%	95%
100	60	61
150	58	59
200	57	58
250	56	57
400	55	55 ½
600	54	54 ½
1,000	53	53 ½

In other words, with a total response of 200, the difference between the two tests would be significant at 95 per cent confidence only if the larger of the two responses were 116 or more. (This method is a form of the chi-squared test, and should be used, instead of the significance-testing method described earlier, whenever you are dealing with a response rate below 0.1 per cent or are testing two unmatched audiences against each other.)

The cost of testing

Testing has a cost – and therefore requires a budget. The cost of testing stems from such items as:

- ✘ Higher print and production costs. Shorter runs and more variants inevitably result in higher unit costs.
- ✘ Loss of discounts. Variations of materials and timing can lead to forfeiture of volume discounts on print, postage and media buying.
- ✘ Executive time. The cost of devising, implementing, controlling and reporting upon a complex test programme is extremely significant and should not be overlooked.
- ✘ Higher proportion of fixed costs. Developing a number of new test approaches can involve additional fixed costs, such as creative origination charges or agency fees.

Your test budget should be seen as an investment in the future, and the less you know about your market, the more you would invest in testing.

It should also be remembered that testing is not necessarily expenditure that cannot be recouped. As part of an ongoing programme, tests may well earn a higher return on investment than the remainder of the programme even in the short term, if the tests have been well chosen and implemented.

Monitoring and evaluating test results

A carefully structured test programme will count for nothing if the results are not properly monitored, analysed and evaluated.

Monitoring and evaluating systems can be divided into five stages as follows:

Stage 1. Information processing

Stage 2. Response summaries

Stage 3. Results analysis and evaluation

Stage 4. Programme reviews and reports

Stage 5. Campaign library maintenance

Stage 1. Information processing

Personnel who physically handle responses must be briefed with relevant campaign details (products, timing and source codes) and should also appreciate the need for accuracy.

The response handler's brief should include:

? When response should be counted (daily or weekly).

? Exactly what should be recorded (single/multiple orders, order values and source codes).

? Checks to be made to ensure source codes are not wrongly captured. The system should recognise current promotion codes and input data should be verified.

Where the campaign is being conducted through digital media, response tracking can be handled automatically – provided that the requirements are thought through and fully specified in advance.

Stage 2. Response summaries

When the first orders or enquiries are received, systems should be in place to provide ongoing response summaries. Although it is dangerous to draw too many conclusions from early results, where response is exceptionally high or low, early warning allows immediate adjustment of fulfilment resources. The following should be considered:

? Will response actuals be shown against response projections?

? Who will perform what analyses and how often?

? Has an analysis schedule been set up?

Stage 3. Results analysis and evaluation

Once campaign results are final (or if final results can be accurately projected), they should be carefully analysed. Results analysis is two-stage:

1. **Initial response analysis** establishes that results are statistically sound, i.e. the absolute number of orders received is adequate and differences statistically significant.

2. **Back-end (subsequent customer performance) analysis**. Once all campaign costs are known and performance rates established (e.g. conversion from enquiries to sales and bad debts), the profitability of the test programme can be calculated. Depending on test objectives, it is often not until this stage that meaningful conclusions can be drawn.

For example 'two incentives' may outpull 'one incentive' at the enquiry stage, but a poor conversion rate may mean that 'one incentive' is the most profitable approach.

Remember, when test results are analysed, additional costs attributable to testing must be taken out of the calculation. Rollout cost projections should be compared with control costs to establish which is the more cost-effective, test variation or control, in the long haul. Higher origination costs, higher costs associated with smaller print runs and discounts etc. need to be deducted to reflect the rollout conditions.

Stage 4. Programme reviews and reports

To avoid some of the problems outlined in this section it is important to set up formal review meetings or, at the very least, structured test programme reports in order to:

- ✔ Confirm the validity of test results.
- ✔ Objectively assess results and evaluate against test objectives.
- ✔ Balance the results with judgement and past experience. Do any external variables or circumstances need to be considered?
- ✔ Agree key decisions affecting future marketing activity and adjust future programmes accordingly, i.e. should the control be changed, should you adopt the low price, do you retest or drop the new incentive?

Stage 5. Campaign library maintenance

A well-maintained campaign library is essential and should provide a permanent record of all key marketing information for future reference. Details of test programme results and objectives, analyses, recommendations and decisions taken should all be accessible, together with appropriate sample packs or advertisements.

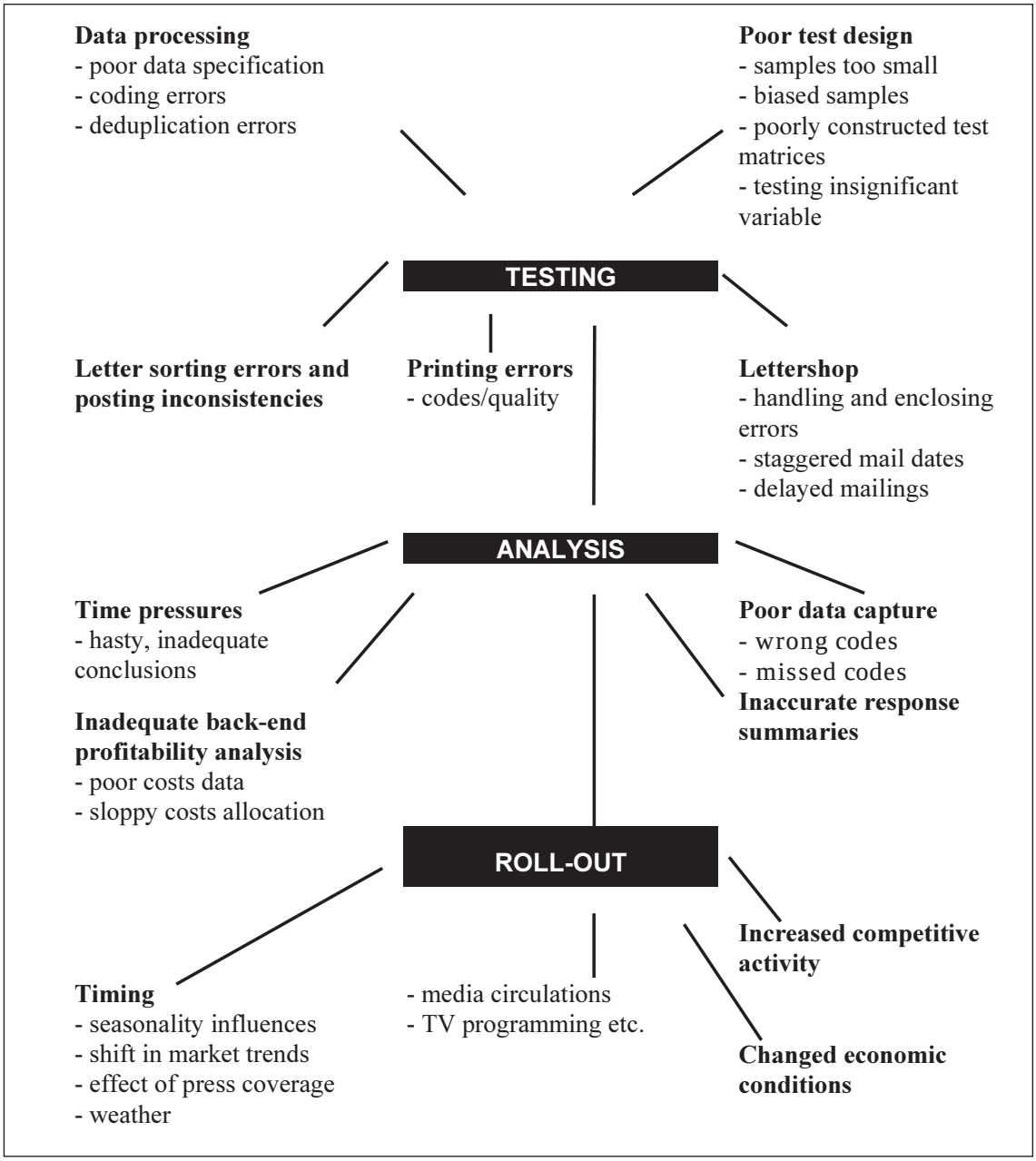
All too often such an information base is not properly maintained or is made the responsibility of a junior member of staff. Having invested time and money in acquiring direct marketing knowledge, you should carefully record, maintain, and refer to it regularly.

How to ensure the reliability of testing

At several points during this section we have stressed that test results should be valid when compared to control all else being equal or, as the scientists say, 'subject to consistent conditions'.

We have dealt with statistical error, but what about operational error and other factors outside our control? The graphic in figure 8.2.2 draws attention to some of the many factors which can lead to predictions being rendered inaccurate and thus to disappointing outcomes:

Figure 8.2.2 **How testing and forecasting can go wrong**



The role of judgement

The theory and principles behind direct response testing are sound. The statistical concepts used are reassuring. However, it must be stressed that testing is an aid to judgement and not a replacement for it. Test-based experience must be modified with judgement in order to reach decisions.

TEST RESULTS + JUDGEMENT = DECISION

Testing can provide invaluable guidance and can dramatically improve long-term profitability, but it is an imperfect science. Many variables outside your control can and do affect test results. It is as well to be as aware of the pitfalls of testing as of the opportunities it affords.

Press testing

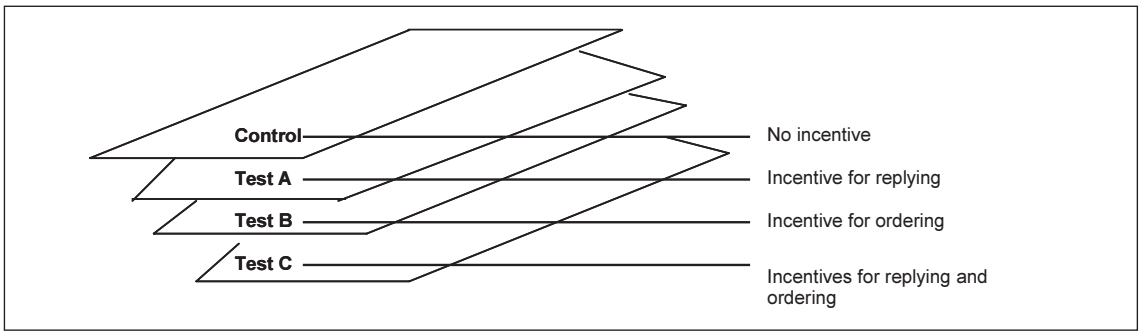
For direct marketers developing customer acquisition programmes across a range of media, testing in space advertising can be as important as direct mail testing.

But, whereas direct mail permits testing a wide range of variables very quickly, space advertising opportunities for testing are fewer. Advertising tests must accommodate the physical limitations of print media. Several techniques are used for testing advertising in space and inserts, as follows:

Perfect splits using inserts

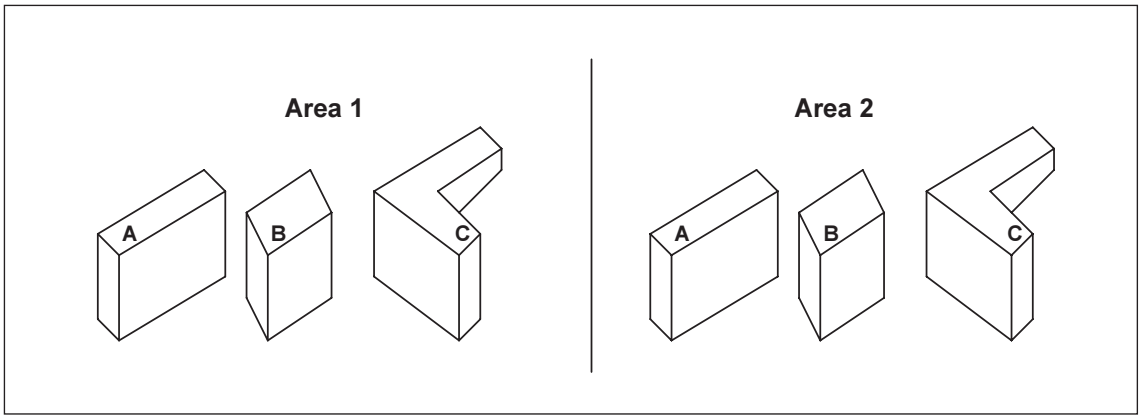
A number of versions of an insert can be interleaved or ganged up at the print stage before being supplied to the publication. In this way a number of options (usually not more than six) can be tested in any one issue of the host medium. This method is ideal for accurate testing of such variables as product, price, incentive, headline, copy and illustrations etc. – anything printed on the surface of the insert (figure 8.2.3):

Figure 8.2.3 **Perfect splits using inserts**



It is more difficult to test different formats of inserts (e.g. unequal sizes, shapes or thicknesses) by this method, since printers and publishers cannot easily arrange random distribution of different items. However, many publishers do offer test facilities for unlike inserts by ensuring that batches of test inserts are delivered to matching areas in equal quantities (figure 8.2.4):

Figure 8.2.4 **Testing formats of inserts**



A further advantage of insert testing is that any number of media can be included on the test schedule with no increase in cost or complexity. Results can be measured not only by test, but also by medium. Interaction between medium and test can also be observed in this way.

NB: Similar procedures can, of course, be applied to door-to-door testing.

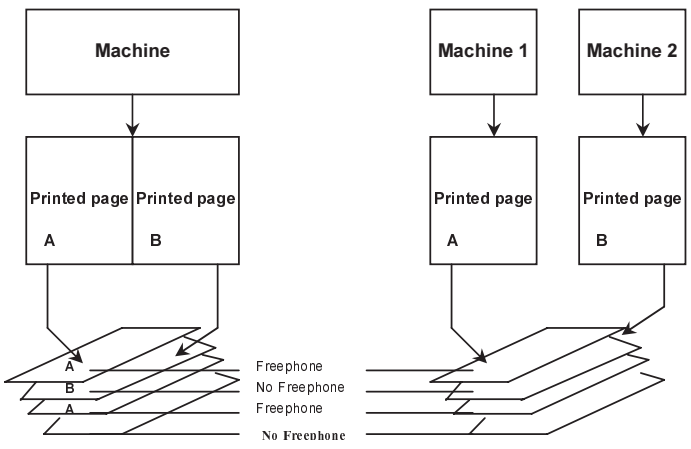
Space advertising: A/B splits

Many publications are printed two-up, meaning the printer produces two copies simultaneously from each machine. By changing a plate affecting half of the press sheet, a perfect 50/50 or A/B split can be achieved. In this way you can arrange for each alternate copy of a publication to contain either advertisement A or advertisement B.

Some publications offer split-run facilities using several machines printing simultaneously. You can arrange for half the machines to print your A copy and half the machine to print your B copy (figure 8.2.5).

NB: Not all publications offer A/B splits: in fact the number that do has declined over time. For up-to-date information on this situation, you should apply to the 'newspaper publishers', or the Periodical Publishers' Association.

Figure 8.2.5 **A/B Splits**



A/B splits can be used to compare a wide variety of variables but always two at a time. In theory, A/B splits are perfect samples, since each alternate copy of the publication contains copy A and copy B retrospectively. Excellent samples of several million copies can be achieved with national newspapers or combinations of magazines.

However, the results of A/B splits, while extremely useful, are not individually reliable. For example, due to breakdowns or other operational reasons, a printer may not always print equal numbers of your A and B advertisements, although this fact will be unknown to you. A/B splits are a facility offered to advertisers at a nominal cost and are generally outside the publisher's legal obligations. Therefore no redress will be available.

A cautionary tale

Some years ago, the agency for Plumb's stretch covers booked a split run in a national newspaper for their client. By mischance, they sent the same copy for each half of the split-run test. The only difference was the key or code number in the coupon on each ad (used to identify the source of responses). One key number pulled 14 per cent more replies than the other key number.

Earlier in this chapter we devoted much space to a discussion of assessing test results. We gave particular attention to assessment of results where the response rate is very low – as is normally the case in space advertising. But whatever the response rate may be, and whatever the statistical methods employed, the golden rule of A/B splits, especially when using newspapers, is not to attach too much credence to a single result. Always repeat important tests several times before drawing far-reaching conclusions, above all if the results are counter-intuitive.

Many newspapers mark their editions with a small 'A' or 'B' at the top or bottom of pages on which split copy occurs so that you can spot A/B splits by acquiring several consecutive copies.

Cross-over testing

Cross-over testing is a technique sometimes used where individual publications do not have A/B split capabilities but where tests are important for the product. Cross-over testing can also be used for testing pairs of advertisements which are different in size, position and use of colour etc. It is, however, less reliable than A/B split testing and far less reliable than testing with inserts.

Using two similar publications, advertisement A is first run in one magazine with advertisement B running in the other. Then, in a subsequent issue, the advertisements are switched. The results are aggregated over the two issues and comparisons drawn, as in table 6.2.10:

Table 8.2.10 **Example of a cross-over test**

	Magazine 1			Magazine 2	
	Copy	Response		Copy	Response
1st issue	A	240		B	120
2nd issue	B	180		A	160
Total		420			280

Clearly, extrapolating results from this cross-over test is not an entirely satisfactory basis for an important decision. As with A/B splits, it is unwise to base decision making on isolated tests but to repeat important tests several times if possible.

Press testing: a summary

When testing a publication, use your control advertisement where possible.

- ✔ If your control advertisement is particularly expensive, experiment with smaller or cheaper spaces.
- ✔ Different media and copy often produce widely varying ongoing customer performance (the so-called 'back end'). It is therefore wise to calculate back-end results carefully before settling on control advertisements and media.
- ✔ Never rely on single press test results. The more critical the decision, the more tests should be carried out (say four to six). Fewer tests than this may yield inexplicably contradictory results.
- ✔ Where two advertisements regularly produce similar, satisfactory results regard this as a good outcome. It could mean you have a second winner!

In summary, testing in press (inserts or space) offers less flexibility than direct mail testing. It is also less reliable in its results. But there is no other way to compare the response-inducing power of advertising in newspapers and magazines. The important point is to be aware of the deficiencies and to retest frequently.

Testing in non-print media

The principles of testing are the same whether the medium is direct mail, press insert, DRTV, email, SDMS or whatever. The test must:

- ✔ Be measurable
- ✔ Have the minimum number of variables at play
- ✔ Have test criteria clearly identified for post-campaign evaluation

However, across the individual media options there are many factors that affect the type of testing and evaluation that is possible.

The internet, on the other hand, has the ability to provide an audit trail to rival all others. The downside is that in the majority of cases the profile of the audience

and impacts can only be assumed by association with the site content, as there is very little site audience research, except where visitors have to carry out a detailed registration process.

Television and DRTV

In drawing up an initial test we need to identify which are the factors that are likely to have the greatest bearing on the overall results, and which can be considered opportunities for further refinement. TV has a robust electronic audience measurement system, BARB, which gives us accurate impact data, audience demographics and programme and advertiser details against which to correlate response information. The aim must be to minimise the capital risk while maximising the response information available for robust analysis.

Cost of entry: what airtime might we use, and what might it cost?

? Who are the core prospects? When do they watch?

? Do we need to chase a specific audience or can we track down low-cost airtime?

? When do the competitors advertise?

? How long do we need to run the schedule to gain a robust insight?

Cost is influenced primarily by the choice of station, the time of year, the daypart and the commercial length. For example, across the ITV stations alone, the regional and monthly variations, combined with the difference in station universe, mean that the capital cost of 100 x 30-second TVRs could vary from as little as £4,000 to as much as £150,000. Cost-per- thousand (CPT) could be as little as £1.50 and as much as £20.

So, to give our test the best chance of success, we need to construct our schedule so that it delivers the required audience at the most cost-efficient price. A schedule constructed from a selection of the many satellite stations is definitely a good route when possible, and using the cheaper months, such as January and February, rather than say, September, is an easy but significant way to reduce initial costs.

Let's work through an example:

Advertising category:	Mail order
Product category:	**Manchester United 1999 Treble Collection**
Price:	£295 inc. VAT, for boxed set of four coins
Payment;	Credit card or cheque, in full or instalments
Proposition:	Commemorative limited edition, 24-carat gold-plated
Media objective:	To test DRTV as a marketing option
Budget:	Not more than £70,000, including media, production and call handling, excluding costs of product
Target returns:	130 per cent of spend (i.e. £100,000), sales of at least 400 boxed sets
Evaluation criteria:	All costs associated with creative and production to be amortised across test campaign

Cost of entry

- ✔ Target audience research shows that the product purchasers from press and direct mail advertising are male biased.
- ✔ A planning consideration should be to investigate whether results are stronger from the more male-orientated stations; however, as the male bias is only marginal, it would be wrong to ignore the female audiences in the test stage. The influence of station profile should be reviewed in the response analysis.
- ✔ BARB audience data can be used to understand the audience profile and assist with the selection of stations for the test schedule.

Table 8.2.11 below shows a snapshot of how BARB data can help us understand the profile of stations. This is from the Sky media site and is updated monthly http://www.sky-skymedia.com/advertising/airtime.

Table 8.2.11 **BARB data extract**

Station	Male	Female	Rank
History Channel	63%	37%	5=
National Geographic	63%	37%	5=
Tara	65%	35%	3=
Sky 1	53%	47%	7
Sky News	65%	35%	3=
Total Sky Sports	70%	30%	2
Total Sky Movies	51%	49%	8
Sky Travel	34%	66%	9
TV	83%	17%	1

- ✔ A review of MMS data indicates that competitors advertise primarily in daytime airtime, across a broad spectrum of stations.
- ✔ The media recommendation is to construct the test across a broad range of stations, utilising a large percentage of daytime airtime, taking advantage of lower-cost audiences.
- ✔ Airtime negotiation resulted in 10+ stations being selected across six contractor groups. Some stations were agreed on the basis of 'CPT agreements' and others on 'spot costs'. The result was four weeks of 60-second airtime, with a regular spread of spots going out.

Response predictions

- ✔ Response calculations are based upon the press results. Response rate increased as TV is likely to generate more 'enquirers'. Conversion to sales decreased correspondingly. 0.01 per cent set as planning target, at 65 per cent conversion.
- ✔ The selected satellite channels consistently show average daily impacts of less than 200,000 or One Adult TVR, generating a maximum predicted spot response of 20 calls per spot. At the same time the stations are sufficiently large to ensure that the response analysis will be statistically reliable.

Results

- ✔ Response rate differed widely by station and ranged from 0.015 to 0.002 per cent.
- ✔ The response rate showed a strong uplift from sports-orientated channels, but not consistently from those with a male bias.
- ✔ ROI varied dramatically by station, ranging from 192 to just 30 per cent.
- ✔ The lowest CPR came from a male-biased station with a CPT index of +90 above the average for the campaign – by no means the cheapest.
- ✔ The two lowest cost stations produced opposing results. The mass entertainment channel delivered a high response rate with low conversion and the other, a more male-biased station, a lower response rate but good conversion to sales and a better CPR.

Conclusion

- ✔ The DRTV test was highly successful on 50 per cent of the selected stations and sufficient data was recorded to ensure that rollout campaigns could be refined to take advantage of the successes.
- ✔ Targeting, at the right price, delivered the best results.
- ✔ The cheapest airtime did not deliver the best results.

Once a product has been tested and it has been proved whether TV can produce viable business results, it is time to look at the evaluation data and isolate the aspects that can be refined.

In our above example, we found that sports-orientated environments produced good quality enquiries, converting to sales at a higher rate than those from more general entertainment environments. Many of the entertainment stations produced excellent response rates, but not of sufficient quality to lead to a sale – which after all is the end-game. The test also proved that the buying CPT was a factor that helped reduce the cost-per-response in some cases, but it was not the biggest determinant of success. In constructing future activity, the media planner can build on these factors, introducing additional stations that match the profile indicators. In addition, the next schedule can be constructed to enable evaluation of other influences, such as times of day, days of the week and break patterns, in order to see how future campaigns can be further improved.

Below we have summarised a checklist of a number of the criteria that can influence the success of DRTV campaigns:

- ✔ Channel and station: ITV regions, Ch4, Ch5 and GMTC macros
- ✔ Regionality: North versus South etc
- ✔ Programmes: genre (e.g. sport versus film), audience profile etc
- ✔ Day of week: weekday versus weekend
- ✔ Time of day
- ✔ Break position: end breaks versus centre breaks etc.
- ✔ Positions in breaks: first ad, last ad or just 'in there somewhere' ad

- ✔ Effects of proximity pairing: commercials close to each other, in the same programme or clock hour, versus more random spacing
- ✔ Short-term and long-term frequency effects over days, weeks and months
- ✔ Creative executions and creative testing are usually longer-term issues than media testing, but the following are common criteria to evaluate if you can: commercial time lengths, use of cut downs, call to action reminders, phone number types – i.e. Freephone versus lo-call, duration of phone number display, voiceover and copy rotation.

Digital TV and iTV

Tests that are currently taking place all require much higher levels of investment than for traditional DRTV, principally to cover the cost of the computer software programming part of the equation.

The scenarios that are possible include:

Having a presence on one of the interactive areas such as Open. Clients using this Portal iTV include Domino's Pizza and Woolworths. To enable the interactivity of this area the viewer needs to plug their set-top box into the phone system, prior to which the viewer is receiving a continuously broadcast, navigable stream of text and video from the open satellite. So for Domino's you can select your pizza topping and flavour of cold drinks, but the order has to be placed 'online'. The open database reads the delivery address and identifies the nearest Domino's franchise, and sends them the order details. The 30-minute guarantee is still maintained and the system appears to work very well.

On the ITVDigital terrestrial platform (formerly OnDigital), CarltonActive clients have tested interactive commercials on Carlton Food Network and Carlton Cinema. Prompted by the on-screen strap, the feature allows viewers to get more involved with the commercial message. A number of major brands have taken part to date:

- ✔ **Max Factor**: make-up tips and product information, plus redeemable coupons for viewers.
- ✔ **Bounty**: the kitchen towel. Details on product range, with tips, celebrity endorsements and a chance to win a holiday to Florida.
- ✔ **M&G**: detailed information on M&G's financial products.
- ✔ **AA**: information and quotes on insurance policies, with special promotional offers

These tests did not take the viewer out of the programme stream but are non-transactional.

Where platforms offer a return path (phone, cable link or 'web on TV' access via ONnet) it is possible to pass captured data from the set-top box to an online server and pass data back to the set-top box from the server for presentation to the viewer. TV banking is a good example.

At the time of writing, testing interactive response is both complicated and expensive. This may change as the services become more widely available, particularly if some of the common operating software is made available to work across the various platforms.

Radio

Radio planning is covered in detail in chapter 5.6; however, as previously, we have detailed below the core aspects that should be included in an initial test matrix.

Like DRTV, the basic currency for radio is audience impacts, measured against a number of different target groups, such as 'all adults', 'men', 'women', ABC1s etc. derived from the RAJAR diary research. Airtime is purchased on the basis of an agreed cost-per-thousand (CPT) for the particular audience. However, pricing is much more stable than is seen with television, with minimal monthly or regional variation.

So core test criteria to include are very similar to DRTV:

? Who is the target audience?

? Which stations deliver the audience profile we need? How much do they cost? What are the impact levels?

? What call handling issues are there?

In response to the above questions you need to consider the following:

✔ The commercial length should be as long as is necessary to put over the sale message and the airtime selected to give the most robust test while minimising capital outlay, about £25,000 to 30,000 should suffice.

✔ Choose two or three test areas. These need to deliver audiences that are as 'average' as possible if you are looking to roll out nationally. If you are a highly regionalised brand, then choose two or three areas that are representative of your general penetration profile. It is important to try and avoid extremes, i.e. the town where you are most successful, or least successful, because this makes predicting rollout response rates more difficult.

✔ Run your commercials in all time bands that the call handling resource permits. Often daytime (ironing and cleaning time) delivers a higher response rate, but it is often beneficial to let the post-campaign statistics tell you this, and ensure you identify the best parameters for your own brand. Other factors that help response efficiency include:

- Utilising airtime at the beginning and end of the week
- Utilising airtime from midday until early evening
- Utilising airtime in conjunction with TV to establish the brand in the listeners' minds through frequency
- Utilising Freephone numbers – although this also increases the number of miscellaneous enquiries and nuisance calls, affecting statistical conversion levels, tying up the call centre unnecessarily and adding to the telephone bill!

✔ Gauging response is not always easy with radio. Consumers find it hard to recall the invisible medium as the source of an offer, and evaluation does not always show it as being as efficient as other media. But radio is an intrusive and influential medium, which has proved itself many times over for local advertisers. Therefore it can be a great help to set up an evaluation matrix that allows you to understand the effect that radio is having upon

other media if you have them running concurrently, particularly print formats such as door- to-door, direct mail and directories.

We have set out below a scenario that aims to do all that … and for less than £40,000:

Eight-cell radio test

1. Scenario:
   - Radio test for advertiser that currently has a mixed media communications campaign including direct mail, DRTV and directories.

2. Test objectives:
   - To test radio as an efficient response medium.
   - To see whether radio advertising can create a climate of awareness and interest within which response to the door-to-door leaflets, direct mail and directories can be accelerated.

3. Campaign timing:
   - Five-week campaign to build awareness and support the door drops.
   - Identify two to three radio areas, the final selection of which should give consideration to the weight of other media in different areas to ensure that all influencing factors are identified.
   - Identify which stations transmit in the areas. Evaluate profile, especially AM versus FM frequencies, and size – is the area big enough to carry out all the tests you want?

4. Campaign weight:
   - Medium weight to ensure individual spot contribution can be measured/monitored.

5. Campaign day and time band considerations:
   - Airtime to run across Monday to Thursday, and Friday morning, due to call centre restrictions.
   - Mid-morning and afternoon time bands upweighted.
   - Morning drive-time peak spots included driving awareness and communicating the message to a wider audience. Evaluation to determine whether there is uplift from this.
   - Evening drive-time spots to remind those who have been exposed to the proposition during the day, but who were unable to respond at the time.

Example solution:

✔ Granada and Anglia selected.

✔ Broadlands and Piccadilly selected: firstly due to profile match; secondly because transmission areas large enough to track DM and directory response at a postcode level and run robust door-to-door test – 40,000 leaflets targeted to each test cell. News-share and Royal Mail to be compared. Response anticipated to be lower from News-share. Need test to validate theory.

✔ Matrix with other media…

Table 8.2.12

Region A				**Region B**			
Radio station A (Broadlands 102 FM)		No radio		Radio station A (Piccadilly 1152 FM)		No radio	
Area 1		**Area 2**		**Area 3**		**Area 4**	
Norwich		Ipswich		Stockport		Preston	
Royal Mail	News-share	Royal Mail	News-share	Royal Mail	News-share	Royal Mail	News-share
50,000	50,000	50,000	50,000	50,000	50,000	50,000	50,000

Such a test matrix will identify:

- ✔ The recorded responses from radio
- ✔ The comparative response rate between two forms of door-to-door delivery
- ✔ Whether the door-to-door campaign is more successful with radio support or not
- ✔ Direct mail delivered to postcode cells both within the radio transmission area and outside can be used to judge the same potential uplift effect
- ✔ Monthly trends in response from directories can be compared to see if there is any change

Online digital media

Many digital marketers argue that their channel is the perfect direct marketing medium. The low cost of campaign set-up, personalisation and contact creates a very favourable economic environment in which the individual channels can thrive. The high level of measurability and accountability really makes digital marketing stand out. Critics would argue there are limitations to even the most fundamental tracking technologies (cookies can be disabled to prevent accurate activity reporting, as can the images that track email open rates, to name but a couple).

But to focus on these failings seems petty given the sheer scale, speed and ease of online marketing reporting…and the ability to track and measure is central to all direct marketing testing. This section seeks to demonstrate that marketers can test more things, more quickly and more easily than in any other area, and these tests can deliver significant business benefit in the short and long term.

Only a few years ago there were not that many online tools to get to grips with – website design was still an evolving science, there was scarcely a critical mass of addresses to make email marketing viable, and search marketing was restricted to tweaking a site's metadata from time to time. Most of the testing was taking place in the graphical advertising space, where banner advertising was dying a slow death trying to replicate the cost-per-thousand business model of traditional media display advertising.

Today's digital marketers have a bewildering array of tools at their disposal and each one demands a high level of understanding just to deliver an acceptable ROI. Many organisations choose to outsource specific activity to specialists, and this presents a dilemma for those with a strong direct marketing culture…how to ensure testing is central to their suppliers' strategy. A recommendation is to choose partners who understand direct marketing and have a passion for testing. There are plenty of players out there, for affiliate marketing, site design, analytics, organic and paid search or email marketing. Challenge them to incorporate testing into their overall strategy and share with them what you have learned in your offline activity. In this way you will be able to enjoy the benefits of testing without the operational (and knowledge transfer) constraints of having to manage it all internally.

There are a huge number of digital marketing techniques that you need to understand. Every one will require a testing strategy to optimise its performance. These are just the techniques you may consider for acquiring customers:

Organic/natural search
Paid search
Affiliate marketing
Banner/display advertising
Solus email list rental
Newsletter email list rental
Blogging
Podcasting
SMS messaging
Viral marketing
Offline URL promotion
Interactive TV
Microsites
Landing pages

There is insufficient space in this Guide to run through testing strategies for all the digital tools and so it is best to focus on a few principles and see how they can be applied to some of the more widely used applications. Referring back to table 8.2.3 we need to ensure that we are following a business process for each of these specific applications, remembering to link our activity back to the key business objectives.

The world wide web

The basic principles already enunciated for testing remain consistent for the web.

Banner ads

Advertising on the web primarily means inserting banner advertisements on other organisations' websites, inciting the viewer to click through to your site. Accordingly, the first purpose of testing is to discover which websites will give you the most cost-effective results in terms of:

Clickthroughs (visits to your website)
Impressions (number of visitors times number of pages viewed)
Registrations (number of visitors submitting some personal details)
Sales

Clearly the most effective sites are likely to be those with some affinity to your own, and those will be the first – subject to cost – that you will want to test. So, if you are selling car insurance, you might go for personal finance sites; you might look at moneyextra.com, This Is Money, Money Supermarket, or at motoring sites – car enthusiasts, or those looking to buy a car – autotrader.co.uk. You can buy a good sized, four-week test campaign with a good breadth of sites to compare and sufficient budget per site to be taken seriously when entering negotiations, for £20,000 to £40,000.

Secondly, you will want to test the effect of different creative treatments – size of banner, use of colour and animation etc. Since a given banner ad, if exposed too often to the same person, quickly becomes boring, not to say irritating, at least three or four different executions should be produced and distributed to each site selected. This will help to reduce the attrition rate as well as start the creative testing process.

Keywords

Thirdly, you should remember that search engines are a further source of visitors to your site. These visitors arrive as a result of searching on keywords submitted by you to the search engine: you may wish to test the effect of different keywords and different search engines.

Tracking

Just as with testing in any other medium, a full evaluation of a test should take into account not merely the immediate response, but also the subsequent behaviour of the respondents. It is precisely to facilitate this that the cookie principle was invented, which allows the website owner to place a small package of data (or cookie) on each visitor's computer which can be retrieved on the next visit, allowing the site owner to identify repeat visitors and, over time, to compile a profile on each. So consider what you can track and what is meaningful to you:

- Number of registrations
- Profile of registrations
- Site traffic and visiting patterns
- Frequency and number of repeat visits
- Time spent on site etc.

Being clear about the objectives is vital on the web, and in understanding how online activity will integrate with offline media.

Website design

Marketers often see the launch of a website as the end of a process rather than the start of a more important one – once a site is live the real work begins! Your new (or your current) website will probably have been subjected to much technical testing to make sure pages load at acceptable speeds, or links work from one page to another. However, these are 'hygiene' issues compared to the huge business opportunities that optimising the site will deliver.

We do not want to concern ourselves with things that have little impact on business performance – changing background colours or altering image resolution is not the kind of testing we need to concern ourselves with. Instead, think of the big issues relating to your site:

- Why do I lose 20 per cent of my visitors on a specific page?
- What would make more people continue their visit?
- How can I get more people into the 'conversion' funnel (a registration screen, signing up for a seminar programme or adding a product to a shopping cart)?
- Why do I lose 30 per cent of my shoppers at the first stage of the four-page check-out process?
- Why do only 20 per cent of people make it through all four pages to submit an order?
- How could I increase conversion rates by 10 per cent at each stage?

With these business issues identified we must try to express them in terms of potential in order to prioritise them. For example, losing 20 per cent of your traffic at one page sounds like a bad situation, but how much more successful would you be if you could reduce this to 10 per cent? If it's significant, you should develop a few hypotheses and build a test programme to improve performance, and then ensure you have appropriate tracking techniques in place to measure the tests.

Some web marketers like to think in terms of 'funnels' – creating checkpoints at certain pages and counting the traffic passing it. If you consider the whole website one large funnel you may have 100 per cent arriving at it but only 2 per cent reaching the check-out pages and 0.25 per cent converting. Looking at the check-out pages you will have 100 per cent at page one and maybe 10 per cent getting to page four. This 'funnel' analysis is easy to set up using web analytics tools and will prove insightful for a wide range of web traffic analyses, not least in suggesting where you should be focusing most of your testing energy and resources.

For simple issues you can probably construct an A/B split test to deliver the business understanding. An example of this would be running a different page layout at a crucial point in the customer journey to see if you can move more people along to the next stage. A more advanced A/B split test would be refining a four-page check-out process into a two-page process…here you are testing one check-out journey against another check-out journey, not just one page against another.

For more complex issues you may need to consider multi-variate testing. Thankfully, we do not have the costly, lengthy and complex operational issues to consider that a direct mail multi-variate test entails (see tables 8.2.4 to 8.2.7). However, we do need to plan carefully all tests and multi-variate tests need extra attention.

Let's go back to that problem page that haemorrhages 20 per cent of your traffic every day. Consider all the factors that may be affecting people moving on to the next stage of the anticipated journey: What are the navigation options? Are they easy to see? Are they lost among other more distracting images? Does the page carry advertising or affiliate links of any kind? What is the call to action to continue? How might different segments of visitor respond to this page?

By considering some of the factors that may be affecting performance it will be possible to come up with alternative page attributes – a larger navigation button, removal of affiliate advertising, moving the navigation button to a more prominent location, different wording on the navigation button and a clearer call to action. You can then feed these independent variables into a testing module and, over time, the software will present alternative versions of each attribute to new visitors and record the impact. Once an individual attribute has been 'optimised' it will look at another one, and then look at all the attributes in combination. After an agreed volume of traffic (at agreed statistical tolerance margins), the programme will have determined the best combination of attributes – your page is optimised! This kind of testing is becoming more common and, as more technology providers offer web-based, low-cost, easy to operate tools, many more direct marketers will be able to drive up website performance.

Driving traffic

Website optimisation is usually concerned with maximising the numbers of people passing through specific points of a site. If you are reviewing your traffic-driving techniques you may be more concerned with delivering a lower cost-per-visit. Your thinking should still follow the same process as outlined in table 8.2.3 but now you will be applying it to the world of reducing acquisition costs.

Affiliate marketing is growing in importance for direct marketers because it offers risk-free acquisition activity – the most common pricing model is one of a cost-per-acquisition. In simple terms this involves a partner website carrying your graphical advertising in the expectation of getting people to click to your site. If they do click through, a piece of tracking technology follows what they do, and if they end up purchasing from you, the affiliate website will take some commission for his contribution … either a flat rate per sale or an agreed percentage of the order value.

In such a risk-free environment you may consider there is little need for testing but even here you should be looking at a whole raft of possible areas for improvement. Firstly, you may consider the advertisements you are using and should test smaller or larger sizes, or different offers, of different creative elements … these small bits of graphical advertising should be treated as off-the-page press ads and subjected to similar testing regimes. Next, you may care to take a longer-term view of customer value – are any partners or any offers delivering more repeat business, or better average order values in the short term? Some of these learnings may be based on campaign analysis but again, using the process-mapping template mentioned earlier you should be able to construct a test programme for affiliate marketing:

Table 8.2.13

Issues	**Possible issue**	**Test options**
Low first order value	Too offer-driven creative	Get affiliate to run split test on current offer and a less price-sensitive alternative
Low traffic volumes	Poor affiliate traffic	Offer higher commission to wider range of affiliate partners and track best volume deliverers
High bad debt rate	Profile of specific affiliate traffic	Test and monitor through new affiliates, maybe at a higher commission rate

This is a test matrix for one element of one of the 'drive to web' tools at your disposal. Considering the complexities of display advertising, natural search, paid search and email list rental you can appreciate that your online activity should be heavily dependent on testing.

Email marketing

Email marketing is quite a contentious medium. The issues of 'permission' and 'opt-in' versus 'opt-out' have overshadowed many of its achievements. The original internet users coined the phrase spamming to describe distribution of unrequested 'junk' email. Here is how the 'eff' describes spamming:

'An opt-out mailing in which the perpetrator involuntarily subscribes thousands of victims, bombards them with mail, and insists that they manually opt out of the list if they don't want it. Though this mimics postal junk-mailing practice, it is entirely foreign to the internet community's standards of behaviour.'

Chapter 12.2 details the current data protection and regulatory issues in the UK and the EU.

Planning email campaigns

The overall considerations for email marketing are essentially the same as for direct mail campaigns. The database, targeting and selection issues that have been described for mailing list work should be applied to email marketing. The biggest single difference between email marketing and direct mail is cost – but it should not be thought that because the media costs of email are so tiny that any less effort is required at the planning, testing and evaluation stages than would be the case in any other medium. Indeed, given the amount of data generated in the course of an email campaign, the reverse may be true.

There are three main sources of email addresses:

1. **Client/company's own email list**

   As a general principle, companies should try and collect the email addresses of all new customers, while setting in motion a plan of updating historic customer records and getting their email addresses onto the database as well. This way they will build up their own email list very quickly (but make sure you follow the self-regulatory guidelines, see chapter 9.1).

2. **Commercial lists (cold or 'opt-in')**

   In addition, many of the mailing list companies are now collecting email addresses along with their other respondent data. Therefore, while the numbers on an e-list are currently much smaller than on a similarly targeted mailing list, it is now possible to email lifestyle clusters such as motor insurance renewals by month. Testing these lists now is definitely worthwhile, because they are going to grow in number very rapidly and the companies that understand how to work them will gain an advantage.

3. Collection of new addresses

Via online advertising campaigns, data capture banners, questionnaires or microsites using competition or games mechanics.

Once data has been collated it needs to be formatted, cleaned and deduped just like any other list.

Handling the response

Whether you are trying to encourage people to go and see your film or to read your newsletter, once you have set yourself up to communicate with your customers by email, you must acknowledge that they may want to reply. This sounds silly, but too many companies think one way, and enquiries coming in from customers and prospects are dealt with in tardy batches or not at all. At the most general level it must be someone's job to keep on top of acknowledging and processing incoming email. There are also sophisticated software packages known as auto responders that can be programmed to carry out certain follow-up routines with a personalised sequence. These can be downloaded and run by the individual company or subcontracted to companies specialising in auto response. (See www.marketinguk.org.)

Tracking the response

Electronic media give you the ability to track an immense amount of information, and integrated techniques can be added to the core message to gain more detail on respondents. This might be done by inviting respondents to visit a unique entry page and calculating the response rate from that. (For instance, film viewers could be given the opportunity to enter a competition based around the film content). Alternatively you can use trackable email broadcasting systems or *ad-tracking* systems.

In the past few years email marketing has matured as a direct marketing channel. Perhaps most importantly, legislation in 2003 has shaped the context within which marketers can speak to customers and prospects. This has been mirrored by a growing receptiveness of consumers and business professionals to email messaging, and house email files continue to grow at great pace. The costs of managing email communications has dropped and sophisticated web-based technologies bring advanced email techniques within reach of all marketers. Indeed, with its relatively low costs of production and high levels of personalisation, email is recognised by many as one of the most powerful direct marketing techniques.

Traditional direct marketers find testing through email offers numerous advantages:

Testing is fast – the testing phase can be as short as a few days, with some companies compressing it down to a few hours for specific kinds of testing. Immediate response and real-time reporting tools make this pace of activity possible.

Roll-out is very fast – once a winning element has been identified this can be broadcast swiftly, (many platforms broadcasting at over one million messages per hour) with none of the production lead times associated with direct mail campaigns.

Testing is cheap – the complexities of constructing test matrices and message design may be the same as offline, but the 'make-ready' and broadcast costs run to a few pence per message.

As the table below shows, email marketers are now able to set up a range of tests to deliver greater performance at every stage of an email marketing campaign:

Table 8.2.14

Issues	Possible issue	Test options
High 'hard bounce' rates	Broadcast platform blacklisted with ISPs	Send test campaigns to problem ISPs from other technology/service providers' platforms
High 'soft bounce' rates	Problem delivering HTML email formats	Test sending text only versions of messages into problem domains
Low open rate	Lack of engagement	Test different/personalised subject lines and new 'from' fields
Low clickthrough rate	No compelling offers	Test pricing, product bundling, incentives for purchase
High unsubscribe rate	People being over-mailed	Test less frequent contact strategy with key segments
Low conversion rates	Too few people seem committed prospects	Test different registration incentives and track conversion rates from each one

Finally, the email marketers achieving the greatest ROI seem to be the ones using 'advanced' email marketing techniques. These will include as a first step segmentation of your database (at the very least into customers and prospects, but should probably involve some transactional analysis). The second technique is to use the dynamic and conditional content capabilities of email marketing to deliver relevant and engaging messages based on individual registration and purchase activity. Finally, marketers should seek to weave their email activity into the recipient's life cycle with your organisation – specific messaging based on date of purchase or likely replenishment/renewal times have proved to deliver higher ROI than 'monthly newsletter' campaigns. However, as with all generalisations, some of this may not be true for your organisation. You should seek to build testing programmes that identify the optimum ROI through segmentation and personalisation, bearing in mind that at some point the effort expended to deliver another personalised segment will not yield a corresponding return. Only by testing will you determine your email marketing ROI boundaries.

Viral marketing

Not to be confused with a computer virus, the term apparently came from the 'hotmail', free email system, because every outgoing email was branded with a link back to 'hotmail', and an advert for their free email service. Since then, viral marketing has quickly become a hot buzzword in e-marketing circles, due to its innately efficient marketing potential. Viral marketing is the least precise of our current selection, but if the idea is good it can be very impactful and cost-efficient.

Many people use it to stimulate awareness and drive response. TV programmers use it to attract viewers, Nike used it to extend its campaign for the 10 kilometre run, and of course companies who specialist in viral marketing such as **p@nlogic** use it. It created a fun game shortly after Mr Prescott had his punch-up, called Splat the MP, and it made it relevant for the whole election period by including figureheads from all the main parties. Marketing UK has created a page where it

collects good examples of viral marketing, which is worth a visit at http://www.marketinguk.org/examplesofviralmarketing.asp

So viral marketing is the creation of an impactful vehicle, usually a game or competition emailed to a small, but targeted base. The marketer then relies on that base forwarding the game etc. onto other like prospects because they think it will be of interest or, usually, amusement. In this way, the marketing message is transmitted to a wide audience at no further cost to the marketer. This is the electronic equivalent to word-of-mouth, and is very good for member-get-member schemes.

Emerging digital techniques

We have reviewed how testing can play a part in delivering incremental gains in some of the more popular digital disciplines. However, one certainty in this space is that there will be more tools to test as time goes on. While advances in technology for traditional direct marketing don't tend to come around too often, you will need to keep abreast of each and every new digital technology, before your competitors do. So a final observation on testing in the digital landscape is this:

> Test the new techniques as they emerge. Find out how they could help your organisation. Play around with some business models and see if they can make money for you. In the past couple of years we have seen RSS feeds, blogging, podcasting and click-to-call all play a part in organisations' marketing plans – have you tested them yet? Better still, test the next new things – before your competitors get there first!