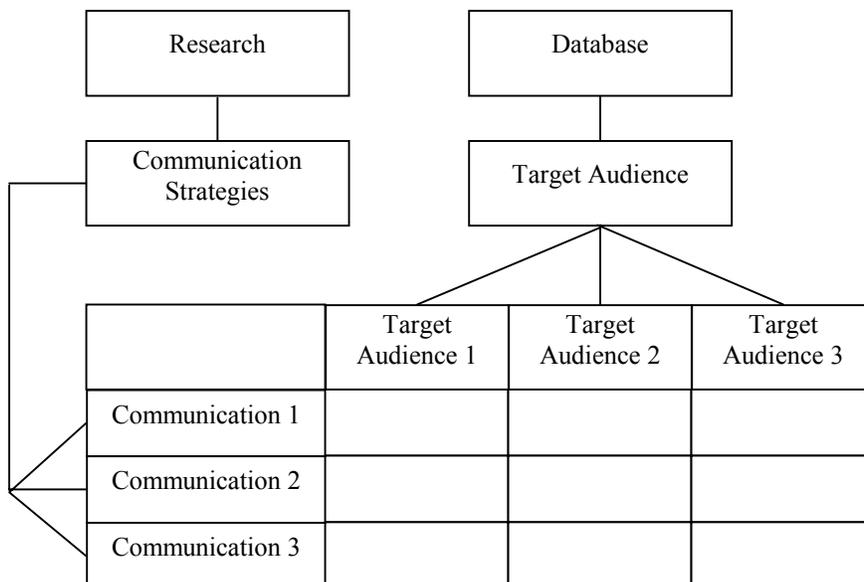




**Using Market Research Information to Impute Actual Customer Behaviour**

The importance of data sources can never be underestimated in any data mining or database marketing project. Of course, the most valuable source of information within these types of projects is prior customer behaviour. External overlay data is also valuable in that we can attach or append demographic attributes whereby the data itself has been aggregated to a postal code level. But one interesting challenge concerning data sources is the use of market research data. Traditionally, market research has been used to understand the underlying attitudes and psychographics of consumers comprising broad target audiences. With this information, marketers have been able to refine their communication and message to their audience but within a mass marketing framework. By that we mean that no targeted lists or segment of names are created for a marketing campaign. With the growth of database/direct marketing, the challenge has been now to integrate the marketing research component within a database marketing program.

In recent years, marketers have used both components in the following manner. With database marketing, marketers are able to identify their specific target lists or segments. Market research is then overlaid around these lists or segments in order to determine the appropriate communication strategies for these specific audiences. Listed below is a schematic diagram indicating of how this integration might occur:



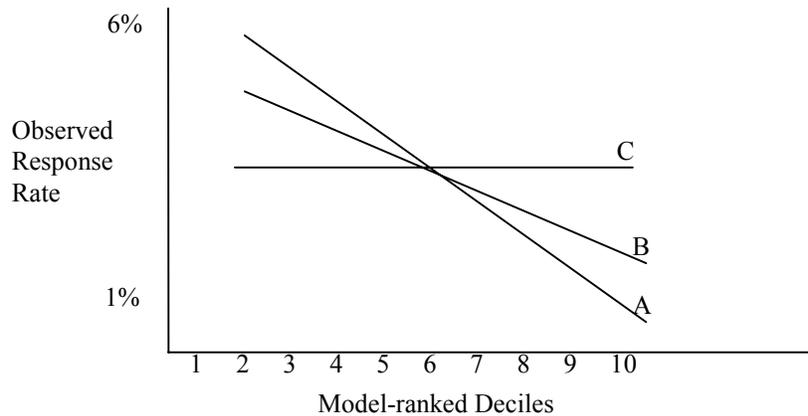
However, it is often the case that database marketers and data miners operate in a less than ideal information environment. In many cases, actual customer information is either very limited or even non-existent. In attempting to analyze customer behaviour within this restricted data environment, the question becomes 'Can we impute some kind of research information as a proxy for customer database behaviour'. As with most questions in database marketing, the typical response is 'It depends.' By this, we mean that the usefulness of research data as a proxy for customer database will depend on both the type of project and the state of the data. The rationale for this kind of uncertainty results from two underlying biases. The first and most important bias is the intended behaviour bias while the second one is the responder bias. Let's talk about the first bias.

The first bias underlines the importance of identifying whether a person who intends to behave a certain way as stated in a research survey actually behaves in that manner as observed within the database. This so-called gap (difference between actual and intended behaviour) is crucial to understanding the degree of accuracy in imputing intended behaviour as a proxy for actual customer behaviour. But like any hypothesis or finding in database marketing, we need to determine what the impact of this gap is and more importantly how to identify an acceptable gap threshold. As database marketers, test scenarios should be created in such a way as to provide measurable results that can provide answers to the above questions.

In our example, we will be using the development of responder models from surveys that are applied against a live customer response campaign. Three surveys will be used in this example, which are identical, but with the following distinct differences:

- A) Actual and Intended Response Behaviour are identical.
- B) Difference between Actual and Intended Response Behaviour is 75%
- C) Difference between Actual and Intended Response Behaviour is 50%.

These three groups have been determined by mapping survey responders with their intended response behaviour back to the database for their actual response behaviour from a live campaign. Responder models are then developed within each of these above scenarios. Each model is developed solely from information off the survey. The model's predicted behaviour is intended response behaviour as indicated from the survey. The survey responder model equation is then applied against a live historical campaign where we can measure actual response behaviour. In this example, each of these survey responder models produces the following actual response rate results when applied against a live marketing campaign.



A: Response rate model developed off survey where actual behaviour=intended behaviour

B: Response rate model developed off survey where the gap between actual behaviour and intended behaviour is 75%.

C Response rate model developed off survey where the gap between actual behaviour and intended behaviour is 50%.

Obviously, the best model is the one with the steepest line or slope, which is Model A. As we can see from the results, there is virtually no lift or rank-ordering in Model C. This would imply that using information from this type of survey with its 50% gap to impute actual behaviour would be a very faulty assumption. Although Model B does produce better results from a rank-ordering standpoint, it is clearly inferior to the results as demonstrated from Model A where the so-called gap is zero. But how do we estimate this impact from a \$ standpoint. One way is to observe the difference in lift between the top decile and the bottom decile.

	Top Decile	Bottom Decile	Lift Ratio of Top Decile to Bottom Decile
<b>Model A</b>	6%	1%	600
<b>Model B</b>	5%	2%	250

Assume that our annual promotional quantity using this information is 100M names whereby the \$ promo cost per name is \$1.00. In order to achieve the same overall results as Model A, Model B would have to promote an additional 140M names. See below for calculation.

$$[100M \times (600/250)] - 100M = 140M$$

At a cost of \$1.00 per piece, the \$ opportunity cost of obtaining survey information with a 0% gap is **\$140M**.

As we can see, the above exercise clearly depicts the impact of this gap in achieving performance. At the same time, this exercise also attempts to put a \$ figure beside this gap.

Although, in this specific case, there is no need to infer intended survey behaviour to actual database behaviour, the intent of this exercise is to use the above results to make judgments concerning similar-type projects with similar type behaviour where the database behaviour is limited or non-existent.

Besides the intended behaviour bias, the second bias is the responder bias. This bias is often overlooked since it is assumed that responders to a survey are representative of the marketing audience. But this is very rarely the case. However, as we saw with the first bias, analysis of data can provide the necessary answers. Insights regarding the degree of this bias can be ascertained by comparing the distribution of key demographics of customers within the survey versus customers within a database.

**Example:**

	<b>Avg. Age</b>	<b># In House</b>	<b>% Male</b>	<b>Avg. Income</b>
<b>Survey</b>	32	1	40%	\$28,000
<b>Customer Database</b>	45	2	52%	\$45,000

The above results would indicate that there is a huge responder bias, which does not reflect the makeup of the general audience (customer database). The learning from the survey being applied towards the customer database would involve insights and recommendations about an audience comprised of single young low middle income women. From these results, it is evident that the results of the survey should not be applied to the customer database. Although these results are painstakingly clear and make our decision a rather easy one in this specific case, in many cases the results will require some judgment and intuition on the part of the marketer in applying survey learning, which may only be slightly biased. In both cases, though, the marketer should be using the results of the data analysis as well as his/her judgment in making the decision to use survey information.

Using market research information in the database environment can provide some unique marketing opportunities. Its effective usage by marketers will only be accomplished with a clear understanding of its limitations. The intelligent use of market research information as a proxy for observed customer behaviour could only augment the marketing decision-making process and provide an additional competitive advantage to its users.