



# Stochastic Solutions

---

Identifying who can be saved  
and who will be driven away  
by retention activity

Using uplift modelling to reduce churn in mobile telephony

# Abstract

It has been repeatedly demonstrated that the very act of trying to 'save' some customers provokes them to leave. This is not hard to understand, for a key targeting criterion is usually estimated churn probability, and this is highly correlated with customer dissatisfaction. Often, it is mainly lethargy that is preventing a dissatisfied customer from actually leaving. Interventions designed with the express purpose of reducing customer loss can provide an opportunity for such dissatisfaction to crystallise, provoking or bringing forward customer departures that might otherwise have been avoided, or at least delayed. This is especially true when intrusive contact mechanisms, such as outbound calling, are employed. Retention programmes can be made more effective and more profitable by switching the emphasis from customers with a high probability of leaving to those likely to react positively to retention activity. This paper discusses how targeting on the basis of such 'savability' can be achieved, illustrating the effectiveness of the approach with case studies. Insofar as a paper can be summarised in a motto, this paper's is "savability is the key to retention activity".

“Savability is  
the key  
to retention”

## Management Implications

- Retention programmes can increase churn as well as reduce it.
- Churn risk alone is not always a good basis for intervention; neither is value-adjusted churn risk.
- The customers at greatest risk of churn are not necessarily the easiest to influence positively.
- Operators not already using control groups in their retention programmes should adopt them as a matter of priority. Properly randomised control groups of adequate size provide the only proven and reliable way of assessing the true impact of retention activity.
- Successful retention programmes (i.e. ones that reduce churn) can often be made more profitable by retargeting on the basis of *savability*. In general, this will remove from the target group a set of customers for whom the programme's impact is negative or marginal, reducing cost and—in some cases—increasing overall retention. It may also add a set of customers for whom the retention activity can be positive, but who are not identified by conventional targeting models.
- Even counterproductive retention programmes (ones that increase overall churn) often have a positive impact for some segments of the population. Retargeting on the basis of *savability* may allow such counterproductive programmes to become effective and profitable.
- Conventional approaches to modelling churn probability are structurally incapable of modelling *savability* directly. However, new modelling methods exist that are capable of predicting customer *savability*; the resulting models are known as *uplift models*.

## 1. Introduction

As mobile phone penetration has increased and market growth slowed, the importance of customer retention has grown. Figure 1 illustrates the evolution of the state of the art, from the initial “land grab” to the increasingly sophisticated targeting on churn risk and customer value that represents mainstream best practice today (fourth approach).

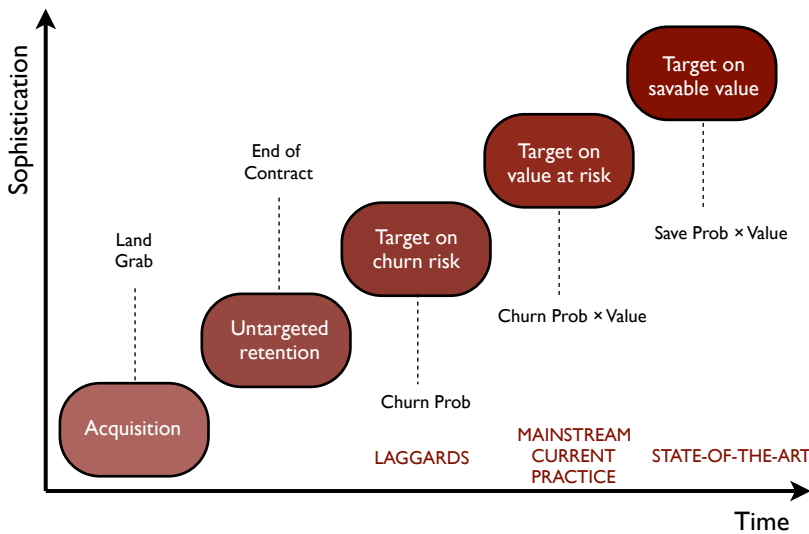


Figure 1: The Evolution of Approaches to Customer Retention

### 1.1. Negative Effects

Accepted best practice for assessing the true impact of marketing actions requires the use of systematically randomised control groups. These provide a statistical baseline against which changes in behaviour resulting from interventions can be measured. As these have been increasingly employed by operators running retention programmes, many have been shocked to observe that some customer segments are adversely affected by the retention activities employed. To be clear: the very act of trying to save some customers provokes them to leave. Far from being a mere statistical aberration, this is a real phenomenon. In the most extreme cases, the overall impact of the retention programme has been found to be counterproductive. There are a number of explanations for this.

“ *The very act of trying to save some customers provokes them to leave.* ”

1. Most customers at high risk of churn are dissatisfied in some way, perhaps because of service, price, coverage or handset. While some such customers will actively approach the operator to terminate, others will not, perhaps as a result of lethargy or resignation. In such cases, active contact with the customer, particularly through an interactive

medium such as a phone call, can act as a catalyst, crystallising a churn event that might otherwise have occurred much later or not at all.

2. Additionally, many customers are antagonised by what they feel to be intrusive contact mechanisms; indeed, we assert without fear of contradiction that only a small proportion of customers are thrilled, on hearing their phone ring, to discover that the caller is their operator. In some cases, particularly for customers who are already unhappy, such perceived intrusions may act not merely as a catalyst but as a constituent cause of churn.
3. Finally, there is also simple awareness of the impending end of contract. While most operators have systems to alert them when customers approach the end of contracts, the customers themselves generally do not. A retention action may inform or remind a customer that he or she has an opportunity to move, and even a fairly satisfied customer may at that point look around for other options. When this happens it is inevitable that a proportion of people will choose to move.

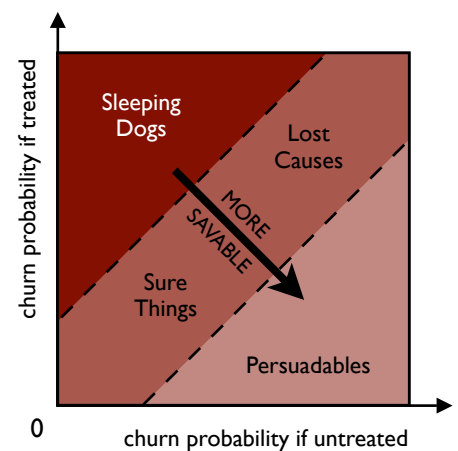
## 1.2. Optimal Targeting

As soon as it is accepted that interventions intended to save customers are a double-edged sword, it becomes obvious that targeting primarily on the basis of predicted churn risk is a dangerous strategy, the more so if the correlation between attrition probability and dissatisfaction is accepted.

The important corollary that this also brings into focus is that even among customers for whom the retention activity is neutral to positive, those at greatest risk of churn are not necessarily those most positively affected by our actions. Indeed, given the correlation between dissatisfaction and attrition probability, there are some grounds for thinking that high-risk customers may be some of the hardest people to influence positively. A segmentation of customers by their different possible reactions to retention activity is shown in the schematic to the right.

The traditional approach is to model the probability of customer churn. As the schematic illustrates, there are in fact two different churn probabilities—the probability of churn without intervention (horizontal axis) and the probability of churn when the customer is subject to the retention activity in question (vertical axis). Which of these an operator models is normally determined by how much retention activity that operator undertakes. If the policy tends towards ‘leaving no customer behind’, it may be that most or all high-risk customers are included in retention programmes and this will tend to mean that the only churn probability that can easily be modelled is the probability of churn when treated, which we denote  $p_c^T$ . (Such models are similar to so-called “response” models, which are often used in marketing.<sup>1</sup>) In contrast, when there is less retention activity, the tendency will be to model churn probability for untreated customers, which we denote  $p_c^U$ . (Such models are more like penetration models.<sup>1</sup>) In practice, the modelling populations are sometimes even mixed. Targeting on the basis of

“ *The customers most at risk are not always the easiest to influence.* ”



- Sleeping Dogs** tend to be driven away by treatment
- Lost Causes** are likely to leave whether treated or not
- Sure Things** are likely to stay whether treated or not
- Persuadables** are made more likely to stay by treatment

<sup>1</sup> Shepard, D. (1999). “Penetration Models vs. Response Models.” Direct. 1st August 1999. Available electronically at [http://directmag.com/mag/marketing\\_penetration\\_models\\_vs/](http://directmag.com/mag/marketing_penetration_models_vs/).

treated churn probability,  $p_c^T$ , tends to focus attention on customers above some horizontal line on the schematic, typically taking in most of the Sleeping Dogs, and Lost Causes, but fewer of the Sure Things and Persuadables. Conversely, targeting on the basis of the untreated churn probability,  $p_c^U$ , focuses attention on customers to the right of some vertical line on the schematic. This is better, in that it will tend to cause more of the Persuadables to be targeted, but it will still capture many Lost Causes and a fair number of Sure Things and Sleeping Dogs.

Current mainstream practice weights these customers by some kind of value metric. This helps in case of high-value Persuadables, but exacerbates the problem with high-value Sleeping Dogs.

Clearly a better approach is to model *savability* (the difference between the treated and untreated churn probabilities), a quantity that increases in the direction of the arrow on the schematic. This allows targeting of the Persuadables without wasting money on the Sure Things and Lost Causes or worse, spending money to drive away the Sleeping Dogs.

## 2. Evidence and Measurement

### 2.1. Two Example Campaigns

We will illustrate the effects we have discussed with reference to two specific cases we have worked on, one from a European mobile operator and another from a US cellular operator. These cases are between them broadly representative of the kinds of results we have seen from successful and unsuccessful campaigns working with a range of mobile operators in Western markets. Key metrics for these two cases are shown in the table.

As can be seen from the figures, the campaign from Operator 1 reduced overall churn from 30% to 25% across the target group, a very positive result. This campaign was very profitable. In contrast, the campaign from Operator 2 actually increased churn from 9% to 10%, a highly undesirable result. Clearly for Operator 2, the first and simplest way to improve things was simply to stop this retention activity. However, the question remained: was it possible that the campaign was actually having beneficial effects for some customers, but that these were being more than offset by negative effects in other segments? For Operator 1, the campaign was successful, so the questions were different—namely, was it necessary to treat everyone? And were there any negative effects being masked by the overall positive impact of the campaign?

It is beyond the scope of this paper to explain how evidence for negative effects can be gathered, and how these can be quantified, or at least bounded, but this has been discussed elsewhere.<sup>2</sup>

Instead, we will simply show what happens when we apply uplift modelling to the problem.

**Key metrics from two Mobile Retention Campaigns**

	Operator 1	Operator 2
Churn rate in control group	30%	9%
Churn rate in treated group	25%	10%
Overall uplift (increase in churn; pc pt)	-5%	1%
Size of control group	10,000	100,000
Size of treated group	20,000	1,000,000
Best targeting cutoff found	75%	30%
Overall uplift at best cutoff (pc pt)	-6%	-1%
Proportion of population required to equal or exceed useful impact of targeting everyone	45%	0%
ARPU	€400	€400
Approximate annual incremental financial impact of adopting uplift per 1m customers	€4,000,000	€8,000,000

<sup>2</sup> Radcliffe, N. J. (2007). "Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models", Direct Marketing Analytics Journal, Direct Marketing Association.

## 2.2. Uplift and Savability

If we want to target on the basis of savability, we need to be able to estimate this quantity for each customer. We cannot measure savability at a customer level because we cannot simultaneously treat and not treat a customer. Estimating savability is a *modelling* problem.

We must first be very clear that conventional churn models do *not* predict savability. A churn model built on a historical population of customers who have been subject to retention activity can be used to estimate the probability that a customer will churn if subjected to that retention action:

$$p_c^T = \text{Prob}(\text{churn} \mid \text{treatment}).$$

Conversely, a churn model built on a historical population of customers who have not been subject to retention activity allows us to estimate the probability that a customer will churn if *not* subjected to the retention activity in question:

$$p_c^U = \text{Prob}(\text{churn} \mid \text{no treatment}).$$

Savability is the difference between these, so that if someone with an estimated churn probability of 3% when treated and 5% when untreated has a savability of 2 percentage points (2pp). We define the uplift U (the net impact of the treatment) as

$$U = p_c^T - p_c^U = \text{Prob}(\text{churn} \mid \text{treatment}) - \text{Prob}(\text{churn} \mid \text{no treatment})$$

so the savability, S, is given by  $S = -U$ . In principle, uplift can be predicted simply by building two models, one on the treated population and one on the control population, and subtracting their predictions; unfortunately, this often works very badly in practice, so novel algorithmic approaches are required. These are beyond the scope of this paper, and some of the techniques are not public, but such methods are discussed in various papers.<sup>3-8</sup> We call models that predict this difference, U, *uplift models*.<sup>2</sup> They are also known variously as incremental models,<sup>5</sup> incremental impact models, net response models, lift models, true lift models,<sup>6</sup> true response models, differential response models<sup>3</sup> and proportional hazards models.<sup>7</sup>

## 2.3. Gains Charts

In order to understand uplift models we first need a means of measuring their effectiveness. We motivate our preferred method with reference to a familiar device—the gains chart.

Gains charts (and gains tables) are among the most common ways of understanding and quantifying the effectiveness of a marketing campaign or a predictive model. Suppose that we have an overall churn rate of 10%. The gains chart then shows the cumulative losses as we run through our customer base, sorted from the highest to the lowest churn scores (i.e. from those the model says are most at risk to those least at risk).

<sup>3</sup> Radcliffe N. J. & Surry, P. D. (1999).

“Differential response analysis: Modeling true response by isolating the effect of a single action.” Proceedings of Credit Scoring and Credit Control VI. Credit Research Centre, University of Edinburgh Management School.

<sup>4</sup> Maxwell Chickering D. & Heckerman, D. (2000). “A decision-theoretic approach to targeted advertising.” Sixteenth Annual Conference on Uncertainty in Artificial Intelligence, (Stanford, CA).

<sup>5</sup> Hansotia B. & Rukstales, B. (2001). “Incremental value modeling.” DMA Research Council Journal, 1–11.

<sup>6</sup> Lo, V. S. Y.. (2002). “The true lift model”. ACM SIGKDD Explorations Newsletter. Vol. 4 No. 2, 78–86. 1

<sup>7</sup> Manahan, C. (2005) “A proportional hazards approach to campaign list selection”. SAS User Group International (SUGI) 30 Proceedings.

<sup>8</sup> Lo, V. S. Y. (2005). “Marketing Data Mining – New Opportunities”. Encyclopedia of Data Warehousing and Mining (ed. J. Wang). Idea Reference Group.

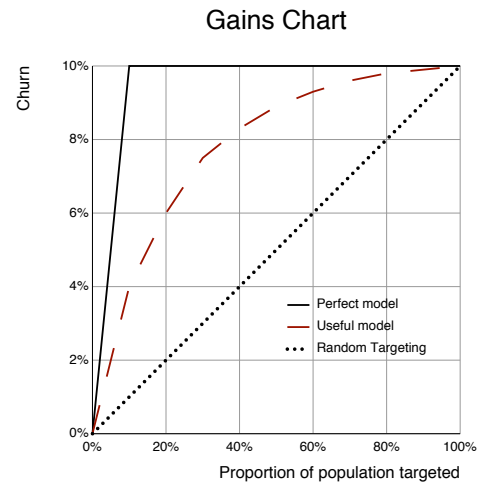
This is illustrated in the graph to the right for three different models.

The solid line shows the perfect model, which gives higher scores to all the 10% of customers who churn than to any of the non-churners. Of course, such a perfect model can never be built in real-world situations.

The dashed line shows a more typical model, characterised by a curve bowed above the diagonal. The customers predicted to churn at a higher rate do indeed churn more (as is indicated by the gradual decrease in slope from left to right) so that this model allows identification, for example, of 20% of customers who account for some 60% of the churn.

The dotted line shows a completely useless model—one that has no ability to discriminate between churners and non-churners.

Thus on a gains chart, the more bowed is the curve above the diagonal, the more powerful is the model. In fact, there is a performance measure, known as the Gini coefficient,<sup>9</sup> that quantifies this as the ratio of the area above the diagonal for an actual model (such as the dashed line) to that for the optimal model (the solid line).

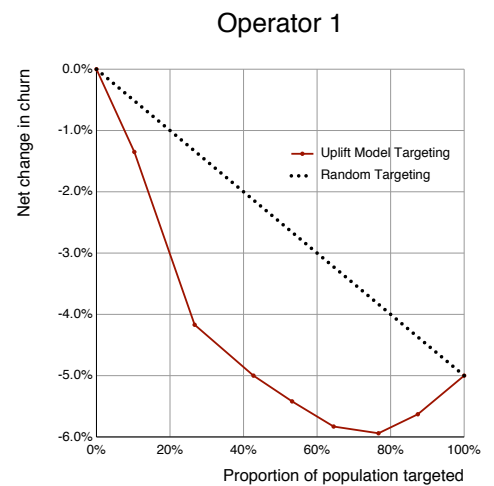


<sup>9</sup> Hand, D. (1997). Construction and Assessment of Classification Rules. John Wiley & Sons (Chichester).

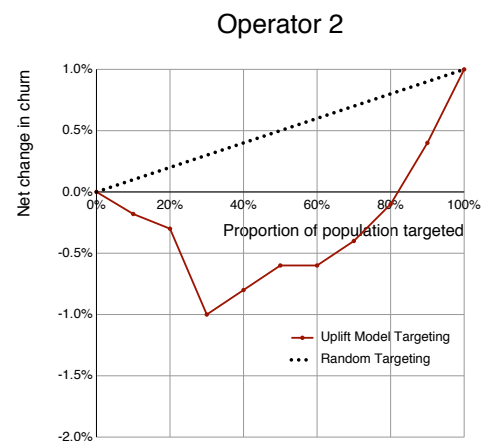
## 2.4. Qini Graphs: Gains Charts for Uplift

Qini graphs and Qini coefficients are generalisations of the Gains Chart and the Gini coefficient to the case in which it is uplift that the model is supposed to predict.<sup>2</sup> The two Qini graphs to the right result from models built on data from the two operators introduced in section 2.1.

The Qini graph is like a gains chart except that now the score used to sort the population along the horizontal axis is interpreted as an uplift score and the vertical axis shows cumulative uplift (usually in percentage points). The Qini graph for Operator 1 ends at -5% because the overall impact of targeting everyone is to reduce churn by 5 percentage points. Similarly, the Qini graph for Operator 2 ends at +1% because the net impact of targeting everyone was to increase churn by 1 percentage point. In both cases, the diagonals show the effect of random targeting, and the solid lines show the impact of targeting using the uplift models.



In the case of Operator 1, the Qini graph shows that by targeting the appropriate 80% of the population, overall churn can be reduced by six percentage points instead of five. (This is *not* 6% of the 80%: it is 6% of the *total* population, i.e. if the right 78% are treated, a net 6% of the total population will be retained who would otherwise churn.) This shows that we can get a double win, increasing retention by one fifth while reducing the contact volume by a similar proportion (c. 22%).



If anything, the results for Operator 2 were even more dramatic. The Qini graph shows that if the appropriate 30% of the population is targeted, overall churn can be reduced by one percentage point. In other words, the retention campaign is quite effective for this 30% of the population. But as treatment continues down the customer file, the negative effects on the next 50% completely

*Qini Graphs are generalisations of Gains Charts to the case of Uplift Models*

wipe out the benefits for the 30%, and targeting the worst 20% increases churn to a level 1 percentage point higher than it would be if no retention activity were carried out at all. Rather dramatically, these figures suggest that the retention activity has a somewhat negative effect for 70% of the population, though this may be misleading: an even better model might exist that would allow more accurate identification of a smaller number of people who are in fact negatively affected.

### 3. Predicting Savability with Uplift Models

#### 3.1. Traditional Churn Modelling

As noted above, there are two kinds of churn model, the difference being whether or not the modelling population has been treated. The modelling procedure is the same for the two cases.

Churn models are normally built by taking a historical sample of the relevant population of customers and using a fitting method (often logistic regression) to relate predictors (“independent variables”) to churn outcome (the “dependent variable”). Here, churn outcome is a 0/1 indicator that might be measured by observing customers for some period such as 6 months. We might find that good predictors of churn include cost per minute of usage, proportion of calls dropped and age of handset. The resulting model might take the form of a scorecard that assigns a number of points for each range of these variables as in the table to the right.

In some cases, it is then necessary to translate the score into a churn probability by applying a simple transformation. In the case of a logistic regression, this would normally be something like

$$p = \frac{1}{1 + e^{-ks}}$$

where  $s$  is the score and  $k$  is a scaling constant. This will map the score into a probability between zero and one.

In constructing the modelling sample, it is of critical importance that an appropriate *observation window* be used. By this we mean that the predictors must measure the state of the customer at some time *before* the modelled outcome, in this case churn. For example, if we use an outcome period of the second half of 2007 (the period during which we record which customers churn), it might be appropriate to measure predictors such as *cost per minute* and *proportion of calls dropped* during the first half of 2007. It is the very essence of all predictive modelling that we build the model by fitting outcomes in the recent past as function of data from the more distant past. We then make predictions by measuring the corresponding variables in the recent past, allowing us to predict outcomes in the future by applying our fitted function—in this case, the scorecard. In doing so, we make the fundamental assumption that the struc-

**A Possible Scorecard for Churn Risk**

	< €0.10	€0.10–€0.25	>€0.25
Cost per minute	-10	+20	+50
Proportion of calls dropped	<2%	2%-4%	>4%
	-20	0	+50
Age of handset	< 1 year	1-3 years	> 3 years
	-10	+50	+10

A person with an average cost per minute for calls of €0.12, 3% of calls dropped and a handset that is 2 years old would get a score of 20 + 0 + 50 = 70.

“ *The fundamental assumption is that the past is a good guide to the future.* ”



tural relationship between predictors and outcomes is relatively stable over time, i.e. that the past is a good guide to the future.

### 3.2. Modelling Savability with Uplift Models

A customer’s savability is defined simply as the reduction in churn probability that results from treatment:

$$S = -U = p_c^U - p_c^T.$$

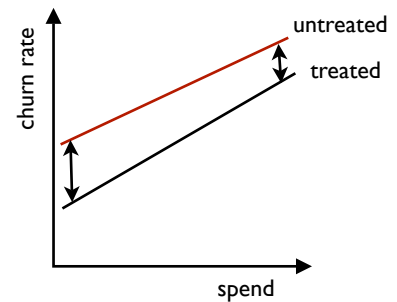
As noted in section 2.2, the complexity in modelling savability arises from the fact that we cannot simultaneously treat and not treat an individual customer. Conventional modelling is based on knowing the outcome to be modelled for each customer in some historical period (in this case, whether or not the customer churned) and then finding correlations between those outcomes and the values of the predictor variables using a fitting or learning procedure. If our goal is to fit the difference in probability that results from our retention activity, conventional regression and similar approaches are of little help, because they all depend on having the known outcome for a historical population to learn from (or regress against).

Faced with this situation, the obvious approach is to build two models, one on the untreated population and the other on the treated population, and then subtract one from the other. In principle, this certainly provides a valid and unbiased estimate of the savability. Unfortunately, however, in many cases it does not work particularly well. We speculate and believe (though it is hard to prove) that there are two main reasons for this.

The first is that in practice the magnitude of the uplift is often small in comparison to the churn rate. For example, for Operator 1, above, the magnitude of the uplift, at 5 percentage points, is around one fifth of the overall churn rates (25% and 30% for treated and untreated populations respectively). For Operator 2, the ratio of uplift to the churn rates is even smaller at around one tenth (1 percentage point against 10% and 9%). This creates a significant problem with “signal-to-noise ratio”, namely that in the main variations in outcome (as fitted by the two separate models) will tend to be much larger than the variations in uplift that we actually wish to estimate.

The second problem is more profound: it is that the goal of the fitting procedure for the two component models (treated and untreated) is not necessarily strongly related to the fitting goal for uplift. For while it is the case that the difference between perfect treated and untreated models would, by definition, estimate uplift perfectly, there is no general reason to suppose that the main drivers of variation in uplift and those of variation in churn will be the same. Indeed, it is not obvious that the factors controlling how likely someone is to leave should bear any particular relation to those governing how that person will respond to a given retention campaign.

So while subtracting two models is certainly a method worth trying, especially when the uplift is large, in general it is both theoretically and empirically better to



In this example, churn increases with spend, but the effectiveness of the intervention decreases (as can be seen from the convergence of the lines). It may be unrealistic to expect a difference of two models to predict this variation in savability reliably.

use a dedicated uplift modelling technique than to model the two populations separately.

In order to apply uplift modelling successfully, just as with conventional modelling, it will be necessary to have a suitable historical sample to allow construction of a valid observation window. The added complication is that now there must be two historical populations, one of which has been subject to the retention activity in question and the other of which has not. In general, the two populations must be statistically equivalent in all respects except the treatment decision, and this is normally achieved by randomly withholding treatment from a random proportion of a target population. In some cases, it is possible to compensate if there is some bias in the allocation of customers to the treated and control groups, but this always significantly complicates the analysis.

## 4. Conclusion

As an increasing number of operators have discovered, retention activity can have negative as well as positive effects. It follows from this that it is dangerous to target retention activity primarily on the basis of estimated churn risk, whether or not this is weighted by customer value. We have shown examples of two retention campaigns, one highly effective and profitable, and the other counter-productive and severely loss making, and demonstrated that both can be radically improved by using uplift modelling to predict savability. In both cases, this has a dramatic and positive impact on the profitability of the campaign, and also on the overall level of retention achieved by the operators. These are not isolated cases, but are, in our experience, rather typical.

The adoption of uplift modelling requires operators to embrace fully the use of systematically randomised control groups, and commits them to using sophisticated, modern uplift modelling methods. These are more complex than traditional methods, but we believe that the demonstrable improvement in results that can be achieved more than justifies such a transition. Once again, *savability is the key to retention activity.*

“ *Both successful and unsuccessful retention campaigns can be dramatically improved by uplift models.* ”

# Authors

## **Nicholas J. Radcliffe**

Nicholas Radcliffe is founder and director of Stochastic Solutions Limited, a consultancy that focuses on customer targeting and optimization problems.

He was previously a founder and Technical Director of Quadstone Limited, where he led the development of their approach to uplift modelling and numerous client engagements.

Nicholas is also a Visiting Professor in the Department of Mathematics and Statistics at University of Edinburgh, where he works in the Operations Research group.

[Nicholas.Radcliffe@StochasticSolutions.com](mailto:Nicholas.Radcliffe@StochasticSolutions.com)

## **Rob Simpson**

Rob Simpson has worked in data mining for business for over 15 years following a PhD in pattern recognition, and has assisted over dozens of major corporations to generate ROI from data analysis.

He has been Head of Customer Services for Portrait Quadstone, formerly Quadstone, for the last eight years, prior to which he was working as an independent consultant.

In recent years he has worked with many European telecommunications companies managing projects including churn, credit scoring and cross-sell.

[Rob.Simpson@PortraitSoftware.com](mailto:Rob.Simpson@PortraitSoftware.com)