Intelligent design: How to model gambler risk assessment by using loyalty tracking data

Tony Schellinck[1,2] & Tracy Schrans[2]

[1]Dalhousie University, Halifax, Nova Scotia, Canada
Email: schell@dal.ca
[2]Focal Research Consultants, Limited Halifax, Nova Scotia, Canada

Abstract

The ability to analyse player data collected from customer loyalty programs, smart cards, and on-line systems by risk for problem gambling has the potential to change the gaming industry and how it operates. Gambling providers are coming under increasing pressure to make use of player tracking data to identify and subsequently help at-risk and/or problem gamblers. Although the prospect of successful identification and intervention is vastly improved by the use of such a system, there are still legitimate concerns surrounding how to implement and evaluate the use of player data for these purposes. To inform ongoing debate, this paper will provide an overview of lessons learned through the authors' work in creating gambler risk assessment models by using loyalty data. This paper has particular relevance for social policy, regulatory oversight, and corporate social responsibility applications.

Introduction

The environment in which gambling providers operate is changing. Jurisdictions are beginning to enforce gaming operators' legal responsibilities for preventing and minimizing harms through legislation, regulation, and licensing conditions (Great Britain Department for Culture, Media and Sport, 2005; Great Britain Gambling Commission, 2005; New Zealand Gambling Commission 2003, 2004). A key feature of these regulatory specifications is a requirement for gambling operators to develop programs or policies for identifying problematic gambling behaviour among patrons. The use of loyalty data and player tracking has been noted by researchers and regulators as an important and reasonable approach to screen for at-risk and/or problem gamblers (Sky City Auckland Entertainment Group, 2007). Growing awareness of potential harms associated with gambling and the subsequent adoption of risk identification and prevention for corporate social responsibility purposes has already led some gambling providers to invest in systems that use player tracking technology for responsible gaming applications (Hancock, Schellinck, & Schrans, 2008). Several gaming operators have also undertaken initiatives to utilize player tracking data

for assessing players' risk levels and mitigating related harms and problems through host intervention programs (Austin & Saskatchewan Gaming Corporation, 2007; Svenska Spel, 2007). Although player tracking data have been used to identify problem gamblers in the past, neither model development nor the standards for evaluating the results and performance of the models have been critically evaluated. From our research experience and expertise in this area of inquiry, we have established a number of criteria that need to be considered in the development of a Gambler Risk Assessment System (GRAS). This paper focuses on specifications for model development by using player loyalty data that describe potential pitfalls and sources of error associated with the process.

The Need for a GRAS

The transtheoretical model of change suggests that a gambler's recognition of their problem is a major step leading toward recovery (Prochaska & DiClemente, 1992). This position has been substantiated by research in the field (Hodgins, 2001; Schellinck & Schrans, 2004a). Nonetheless, it appears that as few as 5% to 15% of problem gamblers seek professional assistance (Schellinck & Schrans, 2004a; Shaffer & Korn, 2002). Moreover, those who present for help are likely the most extreme cases and may have been referred for treatment by other agencies. The remaining problem gamblers may be categorized as (a) seeking assistance informally from a spouse or others, (b) having low motivation to seek assistance despite experiencing harm as a result of their gambling activity, or (c) unaware that they are at risk. In particular, the latter group may benefit from an opportunity to gain insight and a better understanding about their current situation vis-à-vis their gambling and risk for problems.

Some forward thinking operators in the gaming industry feel responsibility for identifying and assisting those who are having problems with their gambling. Indeed, operators may face lawsuits and subsequent financial penalties if found guilty of avoiding a "duty of care" for their customers. As a result, lack of remedial action on the part of a problem gambler or at-risk patron is of concern to the gambling provider, thereby stimulating industry interest in other ancillary methods for supporting identification and support for problem gamblers in non-clinical situations such as the gaming environment. Screens such as the Canadian Problem Gambling Index (CPGI; Ferris & Wynne, 2001), the Victoria Gambling Screen (McMillen, Marshall, Wenzel, & Ahmed, 2004), and the EIGHT screen (Early Intervention Gambling Health Test; Sullivan, 1999), which are largely based upon the South Oaks Gambling Screen (SOGS; Lesieur & Blume, 1987) and the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (4th edition, American Psychiatric Association, 1994), are currently used in a number of venues to assist staff and players in self-identification of risk for problem gambling.

The identification of problem gamblers "on the floor," based upon behaviours or physiological and emotional symptoms, has also proven to be a useful tool (Allcock et al., 2002; Schellinck & Schrans, 2004b). Schellinck and Schrans (2004b) assessed both clearly observable and less visible cues to determine the potential for using these signals to identify problem gamblers in situ. The former behaviours included kicking the machine, getting more cash from automated teller machines, and continuing to gamble until the venue closed.

Examples of quasi-visible cues consist of players' indications of nausea or anger while gambling. By combining cues, the authors were able to identify 86% of problem gamblers with a 94% confidence level, thus corroborating the value of using multiple cues for problem gambling identification. Delfabbro, Osborn, Nevile, Skelt, and McMillen (2007) validated these results in an independent on-site study and also recommended the use of cue analysis for identifying problem gamblers.

Despite the value of using combined cues to identify problem or at-risk gamblers, there are practical limitations to the effectiveness of an "in-house" program that relies solely upon staff observation to meet this objective. To be effective, player information must be collected and combined over time. Site staff may not have the continuity (e.g., have been present to observe all play sessions for a particular patron) or capacity (e.g., ability to observe, store, and assess behaviour for multiple patrons). Staff are often required to perform multiple duties beyond those of gambling customer service. Moreover, staff may be biased in the identification process and the process itself may cause tension for staff and patrons. For example, there may be embarrassment, hesitancy, and/or fear of altercations for staff in approaching familiar or unfamiliar customers to discuss a potential gambling problem. An employee may also be reluctant to identify someone who has been "a big tipper". These various problems can be minimized and/or eliminated by reducing reliance on subjective staff recognition and instead using a more objective method of identification to trigger the process; information (i.e., player tracking data) already collected and stored by customer loyalty programs can be used to produce consistent uniform alerts that systematically direct limited staff resources to where they are likely to do the most good.

The Use of Gambler Tracking Data Based on Loyalty Programs

Loyalty programs, similar to those used by retailers to develop customer relationship plans, have been introduced by casinos across the world. In these programs, play behaviour is recorded when gamblers insert their card into the machine or present it to a table attendant, thus making them eligible to receive bonus rewards or other member benefits. By collecting and analysing behavioural cues that have been measured in an unbiased manner over time, such a database can be used to develop a model to identify at-risk and problem gamblers. The development of a loyalty-based model can also provide a measure for comparison with an "on the floor" observation program. Using two such methods of identification provides a means of confirming and/or validating the accuracy of each and improves the likelihood of achieving successful, targeted outcomes.

There are several drawbacks to exclusively using loyalty data to identify at-risk and problem gamblers. For example, one cannot talk to the gambler to explore the underlying reasons for the behavioural patterns identified in the data. In contrast, working directly with a patron provides the advantage of being able to ask about motivation for a particular behaviour (e.g., "When you gambled, did you go back another day to try to win back the money you lost?"). Loyalty data can identify the days on which an individual lost large amounts of money and can determine if the gambler returned within the next day or so to gamble again. However, these data cannot tell us what the motive is for returning the next day (e.g., Is it to win back losses, or is it simply to continue gambling?). If a person is found to continually return

to gamble the next day after a loss, does this occur because the person gambles only on Fridays and Saturdays each week? If so, regardless of any loss on Friday, the event will be followed by gambling on Saturday. To overcome these problems, we have developed several different measures that collectively serve to capture various behavioural patterns indicative of chasing behaviour. These include defining a loss as a fixed amount (e.g., $200 or more) and as a multiple factor of average losses over the previous loss sessions. The measure also includes variables to capture "returning to play within the next day or so." The new variable is then tested to determine its ability to predict problem gambling, as defined by a standard problem gambling screen such as the Problem Gambling Severity Index (PGSI) of the CPGI (Ferris & Wynne, 2001). Only those variables found to be significant are included in our model development phase. Even though a variable such as chasing behaviour is found to be strongly associated with problem gambling, it does not mean that by itself it can accurately categorize someone as a problem gambler (i.e., with less than 10% false positives). In fact, although most problem gamblers display some degree of chasing behaviour, not everyone who exhibits chasing behaviour is a problem gambler. Therefore, a "chasing" variable is used in combination with other behavioural cues in order to make a prediction that meets the standards for accuracy. For example, including behaviours such as "playing for an average of more than 6 hours per session" and "betting at maximum bet levels on the machines most of the time" may allow us to accurately categorize a player as an at-risk or problem gambler.

Loyalty data on its own does not include information that identifies a player as being in a category of risk for problem gambling. Therefore, a necessary step in using loyalty data to generate an accurate model is to obtain a measure of risk for a representative sub-sample of loyalty players; a random sample of suitable loyalty club members is surveyed and a problem gambling screen is administered. Part of this sample is used to develop the model and the remainder is reserved for the holdout sample. The model is assumed to be valid if it accurately classifies a holdout sample of gamblers into the same categories as does the problem gambling screen used in the survey.

The accuracy of the model is likely to deteriorate over time as the gambler's environment changes (e.g., with the introduction of new games, new forms of gambling, and changed gambling limits). This means that models have a "shelf life" and must be recalibrated periodically to remain relevant for current players. It is difficult to determine the reliability of the model over time, though it is generally suggested that a model of this type is accurate for 2 to 5 years (Berry & Linoff, 2004). Regardless, although updating the model adds cost to the process, it is necessary to periodically ensure that the model continues to perform as indicated and expected by using a new sample of gamblers. The authors recalibrated one casino model that was still quite accurate in its ability to classify members as at-risk or problem gamblers, yet over a 4-year period also exhibited increased sensitivity. This meant that the old model was identifying more of the at-risk or problem gamblers but with a higher rate of false positives (i.e., while the model continued to accurately classify actual at-risk and problem gamblers, more non-problem gamblers were also being flagged or picked up by the model).

Given that model sensitivity and accuracy is so strongly related to the data available in the database and the environmental conditions influencing that data, there are likely benefits from some level of customization to reflect the specific characteristics of the market of interest, pre-empting the concept of a "one-size-fits-all" approach to model development. This means that items that are highly predictive of problem gambling in one market or gaming culture may not necessarily be equally predictive in another.

Another potential difficulty with using loyalty tracking data that is often mentioned by gambling providers (to the authors) comes from the inability to measure gambling at other venues and for other forms of gambling where tracking data are not available. The assumptions are that the problem gambling may be originating or occurring elsewhere (e.g., problems with electronic gambling machines outside of a casino, such as video lottery terminals or Pokies), or that there will be insufficient gambling data gathered by the system to accurately classify a patron. In part, this can be addressed by focusing on the behaviour of regular patrons for a particular operator. First, a general gambling screen is usually adapted to measure gambling problems associated with a specific form of gambling. During the customer surveys used to administer the adapted problem gambling screen, we find that the majority of regular local casino gamblers tend to be loyal to a particular venue and that this activity accounts for most of their gambling expenditures (time and money). Although it is not possible to detect problem gambling for all forms of gambling solely from player loyalty data, if we have sufficient data in terms of gambling activity at the specific site of interest, we can correctly identify and classify a large proportion of problem gamblers among the regular clientele independently of where else they may gamble.

And finally, unless the tracking system is set up in such a way that loyalty members cannot share their cards or that card sharing is minimal, the data cannot be used for modelling. Moreover, gamblers may only use one card at a time; if they lose a card or obtain a new card, this information needs to be connected to their existing play behaviour information if it occurred within the time frame of the model. Otherwise, collected data will be unreliable. There are numerous ways that casinos currently discourage card sharing and multiple card use among members that work equally well for ensuring that loyalty data are suitable for model development (e.g., a card must be inserted during play, reward/points are non-transferrable and only eligible for one card per player, players are periodically rewarded for card use). As biometrics and other portable player identification devices are adapted for gambling applications, this problem is likely to diminish.

Specifications of GRAS

From our experience working with gambling databases over the last 8 years, creating algorithms for commercial GRASs and related applications, and our ongoing research into gambling behaviour, we have derived a list of specifications that need to be considered when designing such a system. The remainder of this paper describes these specifications in detail. The discussion is aimed at those who may not have specific expertise in or knowledge of data mining and modelling techniques but are in a position to evaluate, advise, implement, and/or oversee such systems.

Table 1

Example of a classification matrix for a model identifying problem gamblers

	Predicted		
Actual	Problem gamblers	Non-problem gamblers	Row total
Problem gamblers	60	40	100
Non-problem gamblers	20	880	900
Column total	80	920	1,000

1. Evaluation of Model Accuracy

A GRAS algorithm produced to classify gamblers into risk categories for problem gambling must meet certain criteria to determine accuracy. Typically, the researcher assesses the value of a predictive model by using a classification matrix. The approach used is illustrated in Table 1. In the following example, we have a sample of 1,000 gamblers, of whom 100 have been classified as problem gamblers and 900 as non-problem gamblers on the basis of a screening process.[1] We run our model, which predicts (classifies) that 80 of the gamblers are problem gamblers and 920 are non-problem gamblers. The matrix now provides us with the measures to estimate the accuracy of the model.

As noted by Peng and So (2002), the common measures used to characterize the effectiveness of the model by means of the results of the classification matrix are sensitivity and specificity.

Sensitivity is defined as the proportion of observations correctly classified as an event. In the current example, the event is whether the individual is a problem gambler; we correctly classified 60 of the 100 problem gamblers, thus producing a sensitivity of 60%. Another way to look at this is to say that we are effective in correctly identifying 60% of the problem gamblers.

Specificity is defined as the proportion of observations correctly classified as a non-event. In this case, 880 of 900 non-problem gamblers were correctly classified, giving us a specificity of 97.8%.

In our view, four other very important measures are needed to assess the value of a model. These are the confidence level, false-positive rate, false-negative rate, and overall accuracy.

The **confidence level** refers to the proportion of those correctly classified by the model as an event. In our example, 80 gamblers were classified by the model as problem gamblers, of whom 60 were correctly classified and actually are problem gamblers according to the screen. This gives us a confidence level of 75% (i.e., 60/80). If we approach someone in a venue that the model has identified as a problem gambler, we would want to be highly confident that the individual is, in fact, a problem gambler.

The **false-positive rate** is the proportion of those identified as an event when they are not an event. In this case, 20 of the 80 gamblers identified as problem gamblers by the model are not scoring as problem gamblers on the screen, which gives us a false-positive rate of 25%.

The **false-negative rate** is the proportion of those identified as a non-event when they are an event. Of the 920 gamblers identified as non-problem gamblers by the model, 40 are problem gamblers according to the screen, and so we would say that we have a false-negative rate of 4.3% (i.e., 40/920).

The **overall accuracy** is the proportion of all gamblers correctly classified. In this case, 60 of the problem gamblers are correctly classified and 880 of the non-problem gamblers are correctly classified. The overall accuracy is therefore 94% (i.e., (60 + 880)/1,000).

It should be clear why it is important to know all of these measures when appraising a model.

First, most models can look good on one or more of these measures. For instance, using the data in our example, if all gamblers were classified as non-problem gamblers, the model's overall accuracy would be 90%, its specificity would be 100%, and its false-negative rate would be 10%, all of which appears to be appropriate. However, the sensitivity would be 0%, we could not calculate a confidence level or a false-positive rate, and the model would be useless for identifying problem gamblers.

Second, when the model is designed, the analyst has a choice of which of the matrix criteria to maximize. Increasing the score on one dimension, however, usually reduces the score on another. The best example of this phenomenon is the trade-off that occurs between sensitivity and the confidence level. For example, when we increase the model's sensitivity, we maximize the proportion of problem gamblers that we will identify, but usually our confidence level will drop; that is, the more individuals we classify as problem gamblers, the more difficult it becomes to do this correctly, resulting in a higher rate of false positives. Some problem gamblers may behave in such a manner that they can be clearly identified, whereas others share many characteristics with non-problem gamblers. When we classify the individuals in this latter group as problem gamblers, we will also pick up and misclassify some non-problem gamblers who have similar characteristics.

The decision as to whether one maximizes sensitivity or confidence depends on how the model output will be used. If the goal is to cost-effectively reach as many problem gamblers as possible, maximizing sensitivity makes sense. If reducing false positives is an issue (as might be the case if one were using the information to initiate interaction with a gambler on the floor), a high confidence rate is desired. If a gambling provider specifies a minimum confidence level of 90%, in order to achieve this confidence objective, the modeller may be forced to reduce sensitivity to 20% (i.e., 20% of problem gamblers will be correctly classified by the model). Therefore, the cost of having a high degree of confidence in the classification process will be a reduction in the proportion of problem gamblers identified: a large proportion, perhaps even the majority of problem gamblers, may not be identified by the model.

Table 2
Sample risk classification

Model score probability of being in target segment	Model category	% of gamblers that fall into each model category	% in each model category that are problem gamblers	% of problem gamblers in each model category
50% - 100%	High risk (red light)	6	90	20
10% - 49%	Medium risk (yellow light)	34	50	60
0% - 9%	Low risk (green light)	60	10	20

2. Correct Approach to Classification

An algorithm produces output used to classify a gambler. Usually the higher the score, the greater the probability (i.e., certainty) that the gambler is a problem gambler. Sometimes it is incorrectly assumed that the higher the score (i.e., probability of belonging to the target segment), the greater the risk faced by the gambler. Thus, the probability continuum is used to assign gamblers to categories representing varying degrees of risk (e.g., problem gamblers and medium-risk, low-risk, and no-risk groups). However, people assigned a "medium" probability of being in the target group do not necessarily have a medium degree of risk. They could be problem gamblers whose gambling behaviour is not distinctive enough for the model to categorize as high risk (e.g., it is too similar to the gambling behaviours of low-risk gamblers). Therefore, it is incorrect and potentially problematic to assume that a "medium" score obtained by using the model means that the gambler is at a medium-risk level.

Table 2 illustrates this point. The model is used to assign each gambler a probability that he or she is a problem gambler. The information is then used (incorrectly) to label the patron as a high-risk (often denoted by a red traffic light symbol), medium-risk (yellow light), or low- or no-risk gambler (green light). These three model groups may respectively make up 6%, 34%, and 60% of the gambler population. Of those in the high-risk category, 90% are problem gamblers. We can therefore say that we have a 90% confidence level that gamblers in the high-risk category are problem gamblers. Similarly, only 10% of those in the low-risk category are problem gamblers, and we can be 90% confident that they are not problem gamblers if they have been placed in this category.

However, caution must be exercised in interpreting these categories as a risk continuum. A majority of problem gamblers will fall into the medium-risk category, to make up 50% of the gamblers in that category. Thus, although they are not at medium risk, they are simply placed in the medium-risk category because they cannot be confidently assigned to the high-risk category by using the available data. Similarly, caution must be exercised when considering the green light category. This is the largest group of gamblers, and even though problem gamblers comprise only 10% of the group, this represents 20% of all problem

gamblers, a proportion that is similar to those assigned to the high-risk category. This means that problem gamblers were equally likely to be assigned to the high- or low-risk categories in this model, with the majority identified as medium risk. Therefore, it would be incorrect and potentially problematic for an operator to assume that it is safe to target those assigned to the medium- or low-risk (green light) category with a campaign to increase their gambling.

3. Algorithm Validation by Using a Holdout Sample

When developing algorithms, the modeller creates two (and sometimes three) samples. The first is called the training sample and this sample of gamblers is used to create the models. Modelling techniques such as regression analysis, decision trees, and neural networks all maximize the ability to predict or classify by using the available information contained in the particular data set used to build the initial model (e.g., training sample); the specific characteristics of the training sample are used to arrive at an optimal model. If the sample profile differs in some way from the population at large, the model will use these variables to predict group membership. For example, if problem gamblers in the training sample were more likely to play on Tuesdays than were non-problem gamblers, the model will also use play on Tuesday as one of its variables to classify the gamblers. However, it may be that in the general gambler population, problem gamblers are no more likely to play on Tuesdays than are non-problem gamblers. The randomly selected training sample just happened to have more problem gamblers playing on Tuesdays, resulting in the significant association between day played and problem gambling.

To help guard against the possibility of developing a skewed or biased model, the analyst creates a holdout sample called a *validation sample*. Assuming "Tuesday play" by the problem gamblers noted in our example was a random anomaly, the holdout or validation sample would not have problem gamblers playing more often on Tuesdays. If that is the case, when the model is applied to the holdout sample, it will no longer predict as well because one of the key variables is no longer appropriate (i.e., no longer predictive of problem gamblers in that sample), just as it would not work if it were applied to the general gambling population of interest.

Typically, the ability of the model to correctly classify gamblers is reduced when applied to the validation sample, although in rare instances, the model may perform better on the validation sample than on the training sample. Regardless, the results of the model when applied to the validation sample are felt to be a better estimate of the true accuracy of the model and should always be the criteria by which a model is judged. Thus, when estimating measures such as the sensitivity and confidence levels for a particular model, we use the classification matrix for the validation sample as opposed to the classification matrix from the training sample.

4. Selection of the Validation Sample

Following from Point 3, the validation sample used should not be based on a self-selected sample unless it is first weighted to reflect the distribution of the original training sam-

Table 3
Classification matrix based on self-administered screens to validate model

Actual	Predicted		
	At-risk gamblers	Non-at-risk gamblers	Row total
At-risk gamblers	70	3	73
Non-at-risk gamblers	8	19	27
Column total	78	22	100

Table 4
Classification matrix based on a random sample of gamblers

Actual	Predicted		
	At-risk gamblers	Non-at-risk gamblers	Row total
At-risk gamblers	38	2	40
Non-at-risk gamblers	18	42	60
Column total	56	44	100

ple. This assumes that the training sample has the same profile as the general gambling population of interest. Using a self-selection process is problematic, as it could create an artificially inflated estimate of the model's accuracy. The following example illustrates this effect by using hypothetical numbers.[2]

First, for the purposes of our example, we assume that if a standard screen were administered to a random sample of a venue's annual customers, approximately 40% would be classified as at-risk or problem gamblers (we will refer to these groups collectively as at-risk gamblers). Second, we assume that when presented with the opportunity to fill in an on-line self-administered problem gambling screen, those who are at risk are more likely to complete the screen, given that the exercise will be more relevant to them. If we estimate that at-risk gamblers are four times more likely to respond, 73% of those who fill out the screen will be at-risk gamblers and the rest (27%) will be not-at-risk gamblers.

A model might achieve the classification rate shown in Table 3.

This classification matrix reports a sensitivity of 96% (70/73), a confidence level of 90% for identifying at-risk gamblers (70/78), and an overall accuracy of 89% (70% + 19%). These appear to be very good statistics for model performance by most standards. However, they are inflated because they are not applied to a representative sample of gamblers. If a random sample of gamblers was used as a validation sample, then we might expect a classification matrix as shown in Table 4.

We assume the sensitivity of the model would remain the same and that 96% ($\approx$38/40) of at-risk gamblers would be identified. We also assume that the not-at-risk gamblers are identified with the same accuracy as before, but that their numbers become larger because in a random sample of gamblers, they make up a larger portion of the population. In this case, the overall accuracy drops to 80% (38% + 42%) and the confidence level drops to 68% (38/56). This means that, overall, twice as many people are misclassified (20% vs. 10%) and that, rather than 1 in 10 gamblers being classified as a false positive, the rate of misclassification in this particular example increases to 1 in 3.

Arguments can be made that our example either underestimates the amount of bias introduced because of self-selection (e.g., sensitivity would also likely be found to drop when applied to a random sample), or overestimates the bias (e.g., the likelihood of at-risk gamblers filling out the screen does not differ strongly from not-at-risk gamblers). Whatever the case, we feel that the example clearly illustrates the potential for bias in such an approach. There are methods of estimating the bias due to self-selection, but these methods would need to be applied and new numbers produced before the accuracy of the model is reported. The bottom line is that figures for model accuracy reported by using this form of validation cannot be compared with more legitimate forms of validation until the potential for bias has been assessed and adequate controls introduced.

5. Selection of the Training Sample

The sample used to develop the model should be representative of the sample to which it is applied. If the model will be used to assess regular gamblers, (e.g., defined as those who gamble at least 12 times a year on a regular monthly basis), then the model needs to be trained on a sample with the same profile. There are a couple of ways in which this might not be the case. The model could be developed on the basis of patrons of one type of venue, of one specific venue, or of a particular jurisdiction and then be applied to other types of venues, other venues not similar to the one used for training, or even venues in other jurisdictions. As mentioned previously, the validation sample may also not represent all regular gamblers if self-selection is used.

Risk measures such as SOGS and the CPGI-PGSI are applied to jurisdictions around the world, and so the question arises as to why data-based algorithms cannot be similarly applied to other jurisdictions. The answer is that screens such as SOGS have indicators of behaviours and negative consequences due to gambling that are fairly universal. Stealing to pay for gambling debts is indicative of high-risk behaviour in all jurisdictions, and, therefore, the statement works well as a measure to determine risk levels for players in most parts of the world. However, gambling behaviour measures used as predictors in statistical models are less transferable. There is a defined relationship (usually isotopic) between the dependent variable (being a problem gambler) and the independent variables (maximum rate of play in a session) such that above a certain value for the dependent variable (e.g., I spend more than \$11.50 per minute), the gambler is likely to be classified as a problem gambler. A model developed in Canada in which the maximum bet per spin on a video lottery terminal is usually around \$2.50 will have a lower value for this variable being associated with problem gambling than is the case in Victoria, Australia, where the

maximum bet on a Pokie machine can be $10.00 per spin. In this case, a maximum spend of $18.75 per minute may be necessary before a gambler is designated as a problem gambler by the model. The same variable may be effective in both jurisdictions; it is the cut-off level designating problem gambling that is likely to differ between jurisdictions and venues. The appropriate level can only be determined through empirical research by using a sample that is representative of the gamblers (and play behaviour) in the jurisdiction where it will be applied.

6. Appropriate Choice of Data Used in Modelling

A minimum amount of information needs to be gathered on a gambler before a model can be developed and used for classification purposes. For development of a risk model, there must be sufficient data points recorded for the gamblers (this could be a minimum of anywhere from 3 to 15 times) so that the key predictor variables can be calculated (i.e., populated sufficiently for modelling). Typical of most consumption processes, the majority of gambling activity is accounted for by a relatively small proportion of those who gamble over a year. The actual number varies by venue and jurisdiction, but it can be expected that as many as 50% of annual patrons will only have between one and five sessions of gambling recorded over a year. If it is determined that a minimum of 10 sessions is required in order to provide enough information for accurate modelling of gamblers, then this proportion could be further reduced to 25% of the yearly gambler population.

The amount of information (i.e., sessions) needed to produce an accurate model can be determined empirically by using a set of criteria. However, once the minimum number of required sessions is set, the proportion of people included in the modelling process can be increased by expanding the time frame for inclusion. That is, rather than relying on 1 year's worth of data, the modeller can use data over a 2-year period, which may effectively double the number of gamblers who are classified by the model. We recommend using gambling behaviour over a 1-year period, as this time frame typically corresponds with the time reference for most risk measures or screens (e.g., past-year gambling involvement) and also coincides with operator and regulatory annual tracking and oversight. If a shorter period is used, then we recommend that the risk instrument administered to the gambler also reflects this same time frame.

A common problem in data mining and modelling is that the use of information collected at a certain point may no longer be valid as time passes. Obvious examples are demographic variables such as income, work status, or family composition that become out of date and deteriorate in value as predictors over time. However, this would also include any other information collected at a single point such as attitudinal variables, recent behaviour and experiences, perceptions or beliefs, and other measures often collected in surveys. Some of these variables can change quickly and frequently (e.g., work status) and their inclusion can cause model accuracy to deteriorate rapidly. Other variables based on psychological measures such as attitudes toward gambling can change even more quickly and should be used with extreme caution as predictors.

Additional demographic and attitudinal variables could contribute to a model's ability to classify gamblers; however, exclusion of such variables is pragmatic and one often born of necessity. To include such variables would require surveys of all new regular gamblers with periodic update surveys (e.g., every 2 years) of the total sample of regular gamblers in order to maintain accurate information in the database. Some casinos or gaming operators have 30,000 or more regular gamblers and the cost of maintaining such a system would be prohibitive. The authors have worked with databases for large retail and consumer organizations (e.g., financial, grocery, and insurance) that started building models, including these forms of variables, an approach that was subsequently abandoned because the costs of maintaining the data were too high and usually the data itself was unreliable because consumers refused to answer many of the critical questions. Models have been developed that achieve very acceptable classification accuracy without inclusion of these non-behavioural variables, and so, in our experience, models that do not rely on variables of this type are preferred.

7. The Responsiveness of the Algorithm

The model utilizes behaviours that occur over a specific period of time, say a year, and then creates variables from this recorded behaviour. These variables are then used to predict risk and to assign the gambler to the appropriate risk category on the basis of the predicted outcome. Twelve months of data may be required to amass enough information to make accurate predictions. However, it is also important that the model be responsive to changes in behaviour that produce gambler movement between risk categories (either increased or decreased risk). If a gambler stops gambling because of self-exclusion or a change in play patterns to those consistent with non-problematic gambling, the model output should reflect this within a reasonably short period, ideally at the point when the improved behaviour is found to persist. However, the algorithm should not be so responsive that it reassigns a person's categorization on the basis of temporary or transient changes in behaviour. Algorithms that rely more on complex play patterns rather than on simple frequency or extent of gambling will more likely have sufficient momentum to minimize the impact of temporary or extraneous changes in behaviour that are unrelated to risk reduction (e.g., temporary breaks in play due to travel, health, or financial constraints).

8. Accuracy Over the Full Range of Gambling Expenditures

Modelling analysts will have an easier time assigning those who exhibit more extreme behaviour to the high-risk categories. Hence, those who spend more and who gamble more frequently will likely be categorized as those who are at risk or problem gamblers. However, the majority of those who are at risk may not fall into the extremes, although it is more difficult to identify a problem gambler who does not exhibit extreme behaviours.

For example, risk for problem gambling tends be higher among those who gamble at the high end; high rollers play more often and spend at higher rates, and, therefore, it is relatively easy to maximize sensitivity (reach of at-risk and problem gamblers) by building a model that includes most of these high rollers in the high-risk category on the basis of spending and

frequency. However, this group of high rollers makes up a small proportion of the regular players, many of whom are not having any problems. Simply targeting high spenders means that most problem gamblers in the low-spending segments will not be identified at the cost of including too many false positives from the high-spending category. Gamblers in the low-spending segments, as well as those experiencing difficulties with their gambling, can be expected to make up the majority of the player base.

Thus, a good model should be able to identify a significant proportion of at-risk and problem gamblers among all spending segments without incurring a high rate of false positives.

9. The Shelf Life of the Algorithm

The legal and social environment in which gambling occurs, the alternative attractions available for entertainment, and the nature of people's preferences and perceptions change over time. Changes in any of these factors can impact a given model's effectiveness. Once developed, models are said to have a shelf life. The gambling provider should specify that the accuracy of the model be reappraised at least every 3 years and updated if necessary. A representative sample of gamblers should be used when the model is tested and recalibrated.

Data miners sometimes claim that their models update themselves over time. However, to update their effectiveness, the models must have up-to-date values for the dependent variable, that is, the current status of the gambler in terms of risk. The source of this updated information can be from self-administered risk screens completed by gamblers either on their own or with assistance (e.g., on-site staff), or from access to gambler results provided to the modeller through treatment professionals. However, in both of these cases, the sample used to update the model is very likely to be biased toward those who are at higher risk for gambling problems. This is the same issue identified in the discussion concerning validation of the model. Use of this biased sample to recalibrate the model over time will lead to a biased model that cannot be applied to the gambler population as a whole, as it is only valid for those who self-identify as having a gambling problem. Because of sampling, the reported accuracy of the model will also likely be inflated, especially when compared with models that use a general gambling population sample to update and calibrate.

Summary

Existing means of identifying at-risk and problem gamblers, such as problem gambling screens and the use of cues exhibited on "the floor" of the venue, are limited in value and can be augmented effectively by using a GRAS that is based on player tracking data. This paper provided a list of specifications that are not exhaustive, but should give gambling providers and jurisdictions that set policy guidelines for such systems a better understanding of the characteristics of an effective GRAS. Above all, the system should provide the users with accuracy, broad coverage of gambler segments, real value in terms of identifying at-risk or problem gamblers in a timely and effective manner, and confidence that the system is valid in its categorization of gamblers over the life of the model. There is considerable

opportunity for the implementation of these systems in jurisdictions and markets worldwide, and there will be continued development of the techniques for creating models that will meet the specifications presented here.

References

Allcock, C., Blaszczynski, A., Dickerson, M., Earl, K., Haw, J., Ladouceur, R., . . . Symond, P. (2002). *Current issues related to identifying the problem gambler in the gambling venue*. Melbourne, Australia: Australian Gaming Council.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Austin, M., and Saskatchewan Gaming Corporation. (July 2007). Responsible gaming: The proactive approach – Integrating responsible gambling into casino environments. Retrieved from http://www.iviewsystems.com/assets/products/iCare_Responsible_GamingWhitepaper_V2.pdf

Berry M., & Linoff, G. (2004). *Data mining techniques for marketing, sales, and customer support*. Canada: John Wiley & Sons.

Delfabbro, P., Osborn, A., Nevile, M., Skelt, L., & McMillen, J. (2007). *Identifying problem gamblers in gambling venues*. Melbourne, Victoria, Australia: Gambling Research Australia, Office of Gaming and Racing, Department of Justice.

Ferris J., & Wynne, H. (2001). The Canadian Problem Gambling Index: Final Report, Ottawa, ON: Canadian Centre on Substance Abuse.

Great Britain, Department for Culture, Media and Sport. (2005). *Code of Practice: Determinations under Paragraphs 4 and 5 of Schedule 9 to the Gambling Act 2005 relating to Large and Small Casinos*. Retrieved from http://www.culture.gov.uk/images/publications/GamblingAct2005CodeofPractice Schedule9LargeandSmallCasinos.pdf

Great Britain Gambling Commission. (2005). *Great Britain Gambling Act 2005*. Retrieved from http://www.england-legislation.hmso.gov.uk/acts/acts2005/en/ukpgaen_20050019_en_1

Hancock, L., Schellinck, T., & Schrans T. (2008). Gambling and corporate social responsibility (CSR): Re-defining industry and state roles on duty of care, host responsibility and risk management. *Journal of Policy and Society*, *27*, 55–68.

Hodgins, D.C. (2001). Processes of changing gambling behaviour. *Addictive Behaviours*, *26*, 121–128.

Lesieur, H.R., & Blume, S.B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry*, *144* (9), 1184–1188.

McMillen, J., Marshall, D., Wenzel, M., & Ahmed, A. (2004). *Validation of the Victorian Gambling Screen*. Melbourne, Victoria: Gambling Research Panel.

New Zealand Gambling Commission. (2003). *New Zealand Gambling Act 2003*. Retrieved from http://www.gamblingcom.govt.nz/GCwebsite.nsf/Files/act0351/$file/act0351.pdf

New Zealand Gambling Commission. (2004). *Gambling (Harm Prevention and Minimisation) Regulations 2004*. Retrieved from http://www.gamblingcom.govt.nz/GCwebsite.nsf/Files/Reg2004276/$file/Reg2004276.pdf

Peng, C.Y.J., & So, T.S.H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, *1* (1), 31–70.

Prochaska, J.O., & DiClemente, C.C. (1992). Stages of change in the modification of problem behaviors. *Progress in Behavior Modification*, *28*, 184–218.

Schellinck, T., & Schrans, T. (2004a). Gaining control: Trends in the processes of change for video lottery terminal gamblers. *International Gambling Studies*, *4* (2), 161–174.

Schellinck, T., & Schrans, T. (2004b). Identifying problem gamblers at the gambling venue: Finding combinations of high confidence indicators. *Gambling Research*, *16*, 8–24.

Shaffer, H.J., & Korn, D.A. (2002). Gambling and related mental disorders: A public health analysis. In *Annual Review of Public Health* (Vol. 23, pp. 171–212). Palo Alto, CA: Annual Reviews, Inc.

SkyCity Entertainment Group. (2007). *SkyCity Auckland Host Responsibility Programme*. Auckland, New Zealand: Author.

Sullivan, S. (1999). GPs take a punt with a brief gambling screen: Development of the early intervention gambling health test (Eight Screen). In A. Blaszczynski et al (Eds.), *Culture and the gambling phenomenon: Proceedings of the 12th annual conference of the National Association for Gambling Studies* (pp. 384–393). Sydney, Australia: National Association for Gambling Studies.

Svenska Spel. (2007, June 29). *Svenska Spel introduces Playscan* [Press release]. Retrieved from http://www.world-lotteries.org/cms/index.php?option=com_content&view=article&id=2838%3Asvenska-spel-introduces-playscan&Itemid=100311&lang=en

****

For correspondence: Tony Schellinck, Dalhousie University and Focal Research Consultants Limited, Halifax, Nova Scotia, Canada, Email: schell@dal.ca.

Dr. Tony Schellinck is the F. C. Manning Chair in Economics and Business at Dalhousie University and a Principal and CEO of Focal Research Consultants Limited. Dr. Schellinck has been publishing articles and authoring texts in the areas of consumer psychology, research methods and gambling for over thirty years. Over the last twenty years, he has conducted numerous industry and government sponsored surveys, focus groups and field experiments examining gambling behaviour, policy and practices. Since 1988 Tony has co-authored 15 large-scale peer-reviewed government gambling studies that have had an international impact for gambling best practices and social policy. Tony was a co-director of the Faculty of Management's Informatics Initiative at Dalhousie University and is responsible for an ongoing research program in the area of data-mining, measurement and constructs development, particularly related to human behaviour and gambling. He also researches and develops new analytic techniques specifically suited for large consumer databases. These diverse areas of interest have recently converged as

Dr. Schellinck and his team at Focal develops new methods of analyzing gambling loyalty data and player tracking for evaluative purposes as well as identification of problem gamblers and other high-risk behaviours.

Tracy Schrans is a Principal and President of Focal Research Consultants. Ms. Schrans has an Honours degree in Social and Experimental Psychology, completing advanced courses in data-mining, data analysis, research design, and project management. She is a Certified Marketing Research Professional (CMRP) and professional moderator. She has conducted gambling research in the public, community health, regulatory and commercial gaming sectors since 1988, with specific expertise in methodology, behaviour measurement, and prevalence. Tracy consults on a national and international basis, working with governments and organizations in Canada, Europe, Australia, and New Zealand. She has published papers in peer reviewed journals, co-authored numerous government reports and is a reviewer for various publications and granting agencies. She was appointed to the Board of Directors of the Nova Scotia Gaming Foundation from July 2005–2009 and is currently involved in social marketing and program evaluation for gambling, player tracking and loyalty data analysis, as well as outcome monitoring and public health program evaluation for Addiction Services including alcohol, tobacco, other drug use and gambling. Tracy and Dr. Tony Schellinck are the first researchers to use gambling tracking data to manage risk (CSR) most recently designing new gambling instruments to identify early and advanced risk as well as harm among adults and youth for prevention and social policy applications.

An early version of portions of this paper were presented in Schellinck, T., Schrans, T., & Yi, Zou, (2009). Informing the Debate: Specifications for an Effective Gambling Risk Assessment System Based on Loyalty Tracking Data. The 6th International Conference on Gaming Industry and Public Welfare, Macao, China, 149 – 186.

---

[1]A typical model uses the output of a risk screen such as SOGS or the CPGI-PGSI, which assigns gamblers to risk categories as a dependent or target variable. The dependent variable is used as a benchmark in designing the model to classify gamblers by risk level. An underlying assumption in this modelling process is that there is no error in the dependent variable. That is, the gamblers are correctly classified by the screen. However, we know that application of these screens to the same samples often results in only a 60% overlap in classification (e.g., Ferris & Wynne, 2001), so that the results depend very much on the choice of screen. The gambling provider must therefore have confidence that the screen utilized is appropriate for the venue setting and the type of gambling (i.e., electronic gambling machines and table games) because the behavioural data used to classify gamblers are based on these forms of gambling.

[2]Note that the distribution of risk for yearly customers, although hypothetical in nature, is consistent with multiple studies of gambler populations that the authors have examined over the last 4 years, but is not representative of any specific market or venue.