Summary Report

A Review of the Remote Gambling Research Interim Report on Phase II

An independent review of the report produced by PricewaterhouseCoopers (PwC) and Responsible Gaming Council of Canada (RGC), commissioned by Gamble Aware (August 30, 2017)

April 2018

Prepared by Focal Research Consultants

7071 Bayers Road, Suite 309
Halifax, Nova Scotia, Canada B3L 2C2

Large gambling data sets and risk detection algorithms may have great value in determining markers of harmful or risky behaviour and subsequently lead to the development of harm reduction methodology. A report commissioned by Gamble Aware and conducted by PricewaterhouseCoopers (PwC) and the Responsible Gaming Council of Canada (RGC) to explore this hypothesis was released on August 30, 2017. PwC/RGC collected excellent information both in its survey of gamblers and the corresponding online player data. They described several innovative applications of these data to identify problem gamblers. At Focal Research Consultants, we have been working with similar data sets from diverse gambling providers in Canada, Europe, United Kingdom and Australasian markets. We have built risk detection algorithms for commercial and research applications for more than 10 years and are uniquely positioned to assess the strengths and weaknesses of this report. In the following review which was both self-funded and non-commissioned, we identify a number of issues with the sampling methodology and analyses of the PwC/RGC report. These concerns could have significant implications for the appropriateness of the results and conclusions as reported. The existing data set can be reanalyzed to address many of the shortfalls identified here. We have organized our assessment of the report as follows: 1) Sampling and Segmentation, 2) Analysis, 3) Conclusions, 4) Recommendations. As much of the information is associated with advanced statistical procedures, we have provided a brief summary for the lay reader at the beginning of most sections.

Issue 1. Sampling and Segmentation

Sampling is a key source of error in almost all research[1] and every study includes some degree of bias[2] including the PwC/RGC study. Our primary concerns center on the low response rate that produces a non-representative, skewed sample. It is suitable for constructing algorithms but should not be used to generalize to remote customers at large. Because it is not known how each operator contributed to the overall sample any descriptive analysis is of little value. The development of the algorithms was based on a subset of the original sample utilizing four of the nine customer segments created. This further reduced the sample size available for analysis. It also meant the derived algorithms would only have proven effectiveness for those in the four customer segments included in the analysis. This begs the question as to how effective the algorithms would be when applied to the full player target population.

1.1 Sampling

More than 160,000 customers were surveyed online for the study; approximately 10,635[3] or 6.5% completed the survey that included among other items, the Problem Gambling Severity Index (PGSI). There is no evidence that the PwC/RGC researchers made a comparison to a random sample. As a result, the data are not suitable for producing generalized descriptions or for speculating about the prevalence of specific behaviours among the gambling population. A sample of this type may be used to build behavioural models and markers but not to profile players. This is an important and critical distinction that should be understood by those reading the PwC/RGC report.

Despite the apparently large sample size (≈10,635) the true power of the sample is determined by the sample sizes used to create and then test the algorithms. These numbers were not provided in the report; however, our calculations lead us to conclude that, depending upon how they deleted cases from the analysis, at most ≈2270 gamblers were sub-divided over the 4 customer segments. These four much smaller samples were then used to derive the algorithms. As a result the final samples were too small (n≈567) to create proper tests samples to ultimately provide estimates of the effectiveness of the algorithms developed.

The analysts did not report on the distribution of the sample or response rates over the four participating operators even though this could have been done anonymously (Operator A, B, C, D; Product Categories 1,2,3). The lack of this information makes any descriptive analysis of the

[1] Sudman, Seymour. (1976). Applied Sampling. New York: Academic Press.
[2] Keppel, G. (1991). Design and analysis: A researcher's handbook. Prentice-Hall, Inc.
[3] In the PWC report two different sample sizes are referenced, 10,651 and 10,635. For consistency we are using the lower value for review purposes.

segments of little value. For example, one operator may have contributed a disproportionate number of respondents, and, thus, any unique characteristics for the products and policies of this operator could have an equally disproportionate impact on results. This could be assessed by testing and comparing the resulting models' performance for each operator's sample to see how well the algorithms performed when generalized over the four operators taking part. A hold-out sample could have been reserved for each of the operator's samples for this analysis. Details regarding the role of a hold-out sample are described in Issue 2.

1.2 Segmentation

The segments were initially formed using almost all the original 10,635 customers taking part in the survey but, in building the models, decisions were made by the researchers to remove new customers (≈18% of the sample), the majority of customers scoring at PGSI 1-7 (57% of sample), as well as those falling in customer segments 5-9 (≈61% of the sample once new customers were removed). The analysts decided to focus on identifying risk for problem gambling only among four higher frequency customer segments created using the argument that the majority of problem gamblers (63%) taking part in this particular survey fell in this category.

The use of this method resulted in 78% of the remaining problem gamblers (n=424) falling in the top four customer segments, and, as a result, the vast majority of those scoring PGSI=0-7 (n=5297) and the remaining 124 problem gamblers (PGSI=8+) fell in categories 5-9 were subsequently excluded from the modelling and testing process. The authors did not report on how many no-problem gamblers (PGSI=0) remained in each segment for comparison to problem gamblers. Consequently, we do not know the base rates of problem gambling in each of the segments in order to assess the degree of lift achieved by the subsequent model.

As a result, the models were built and validated using a minority of the total customer sample. The problem gamblers in the high intensity cluster exhibited more extreme intensity of play than the remaining no-problem gamblers left in the group; all the other higher intensity non-problem players were likely removed after the cluster was formed. This would falsely increase the reported predictive value of intensity variables in identifying problem gamblers.

Building the models in such a manner, using a non-representative sample, is only defensible if limited to exploratory purposes. The problem arises when the researchers failed to test these applications properly against a complete and representative test sample of eligible customers in terms of evaluating the resulting model's performance. The results from these narrow segments could not be appropriately extended to inform general policy decisions and best practices across the wider UK remote gambling market.

Issue 2. Analysis

We identified a number of analytical weaknesses in the report that lead to inappropriate conclusions regarding the efficacy of the model in this study. 1) A test sample was not used to assess model performance. 2) Customers scoring PGSI 1-7 were excluded from marker testing and lower intensity problem gamblers in customer segments 5-9 were excluded from the analysis. 3) The report only presents the algorithm's performance statistics for a single period. Consequently, this overstates the precision of the model when compared to running the algorithm over time (e.g., 6-12 months). 4) Data was available for testing their hypothesis that algorithms may not identify multi-site users who are problem gamblers.

2.1 No utilization of a test sample to measure performance

A test sample was not used to measure performance. The PwC/RGC analysts used the k-fold cross-validation technique. This approach may derive the most effective algorithm, particularly when the sample is small. However, there is a distinct difference between deriving an optimal model and testing for the veracity of that model. As the entire sample was used to derive the optimal algorithm using k-fold cross-validation, there was an inherent bias in the process of creating an algorithm that captures the unique characteristics of that sample. Thus, this approach does not eliminate the need for a separate test sample to ultimately determine the performance of the final algorithm.

In this report, despite the large initial sample, the resulting sample sizes for the customer segments were considerably reduced as noted in Section 1.1 of this critique and may have influenced the analysts' decision to forego the use of a test sample. Unfortunately, the validation test sample obtained from a single operator (pp. 63-64) was not representative of the original sample and did not use the PGSI to classify players as problem gamblers. Therefore, it was not of any value in validating the effectiveness of the algorithms created for this report. The authors recognized this flaw in the analysis and called for a more rigorous and suitable validation of their results.

2.2 **Exclusion of PGSI 1-7 customers from testing and the use of methodology based on the skewed sample distribution** (e.g., removal of lower intensity PGs in customer segments 5-9)

Separating a sample into two extreme groups such as no-risk gamblers (PGSI=0) and problem gamblers (PGSI=8+) is a method that has been used previously for deriving algorithms, including in work commissioned by Gamble Aware.[4] In this instance, to obtain accurate performance metrics, the analyst must then apply the algorithm to a complete sample representing those customers it will potentially be applied against. This would include the gamblers scoring 1-7; however, they were inappropriately excluded. Rather, the analysts compared approximately 1032 No-Risk gamblers (PGSI=0) to 424 problem gamblers (PGSI=8+) and then used this same sample to produce their performance metrics.

Using the figures presented in Appendix 10.1 of the PwC/RGC Report we were able to calculate that a precision of 80% reported when using a sample of 'no-risk gamblers' and 'problem gamblers' could drop to a precision of 32% if the full range of customers scoring 0- 8+ had been included in a test sample. Accuracy and hit-rates would also have declined substantially.

In addition, the models were built using self-reported assessments of problem gambling; some respondents may have understated the degree of problems experienced and some may have overstated the impacts; in the case of the current sample all of those overstating would be included (e.g., scoring 8+ on the PGSI) but most of those underestimating their problems (e.g., PGSI scores=1-7) would be eliminated. Further discussion about the implications of this observation is found in Section 4.5.

2.3 Evaluation of model performance restricted to a single measurement in time.

The performance of the model was not assessed over time. When a model is deployed many of the problem gamblers identified each month will be the same (if it is a good algorithm). The false positives, however, tend to be different each month. As the year progresses the number of new true positives, i.e., persistent problem gamblers increases more slowly than the number of those falsely identified. After a year the number of false positives will grow such that the precision, based on a year's application, will drop substantially over a single point-in-time measure. The analysts did not provide estimates of how accurate and effective the models would be over a long period of time, such as a year.

2.4 Assumptions based on number of locations of play.

The rationale for excluding customers based on multiple site use was very confusing and not well-supported (pp. 53-54). The authors stated that the algorithms may not identify multi-site users who are problem gamblers yet they did not test this hypothesis. Given the data was available, they could have determined if their algorithms were equally effective with those identified as problem gamblers regardless of how many sites at which they gambled. From a

[4] Responsible Gaming Trust, Report 3: Using industry data to identify gambling related harm. Authors: D. Excell, G. Bobashev, D. Gonzalez-Ordonez, H. Wardle, T. Whitehead, H. Wardle, R. J. Morris, P. Ruddle. November 28, 2014.

practical perspective a specific operator would not have access to data from other gambling sites, and, therefore, it is not reasonable to recommend adoption of markers based on such strategies if better models could be developed using the available data for each operator.

Issue 3. Interpretation of Results

The primary interpretation weaknesses centered on 1) The use of overall accuracy or classification rate to assess model accuracy; 2) The assumption that the current markers are sufficient for identifying risk); 3) The assumption that those who are not identified by the model are not at risk; 4) The assumption that there are a set of universal markers that can be used similarly across operators, game types and seasons to uniformly identify risk; 5) The practical value of 'Daily Triggers' to detect in-play risk;, 6) The conceptualization of graded intervention.

3.1 The use of overall accuracy or 'classification rate' to assess model accuracy

Throughout the report the authors reported 'accuracy' as one of the performance metrics. The authors defined accuracy as "the proportion of problem and no-problem gamblers predicted correctly" (page 24). This metric actually refers to 'overall accuracy', sometimes called the 'classification rate', which in the case of the PwC/RGC study indicates the percent of the gamblers correctly classified as either problem gamblers (PGSI=8+) or no-problem gamblers (PGSI=0). This metric is not very useful as a measure of a model's performance. For example, if the model classified everyone as a non-problem gambler and did not detect even a single problem gambler, it would achieve an overall accuracy of 93.7%. Such classification obviously has no practical value as all of the problem gamblers will be missed.

The most meaningful metrics are those related to how well the model performs in detecting the desired customer target, in this case 'hit-rate' (percent of target group detected by the model) which was reported, and precision (percent of those identified that actually fall in the target group) which was not reported.

3.2 The assumption that the current markers are sufficient for identifying risk

It is not known if the harm indicators selected were sufficient to detect problem gamblers. The resulting algorithms were not tested on a representative sample of active players so their true effectiveness is not known. As well, the authors relied on an *a priori* theory set in Phase I about what distinguishes problem and non-problem play. This resulted in most of the markers used for algorithm development to be related to frequency and intensity. As a result, the three most predictive variables in the algorithms are intensity based. It is highly likely that if they had created a reasonably large set of behavioural based variables that they would have identified more problem gamblers, particularly those who gamble at less intense levels, with greater accuracy. We would argue that these algorithms are not sufficient until the algorithms are optimized and tested within each operator's environment. It should be noted that the PwC/RGC analysts could use their data to check on the assumptions made regarding the harm indicators selected but did not do so.

Moreover, it is not clear as to the degree to which such segments could be replicated and then used reliably when applied to each operator's customer base. We found no evidence that the customer segmentation clusters were tested for each of the operators as no mention of such an analysis was reported in the PwC/RGC paper. Again, such testing could be undertaken and reported anonymously to ensure operator confidentiality.

3.3 The assumption that those who are not identified by the model are not at risk

The authors also implied that those customers who were not classified by the algorithms as problem gamblers (e.g., with a risk score not exceeding the top 20% threshold) were assumed to be no-risk players.

One example of such an assumption is provided on Page 62 when the PwC/RGC authors introduced the concept of 'Low Risk', 'Moderate Risk', and 'High Risk' based on players' risk scores produced by the model.

This is not a correct assessment of risk. Because a person has a lower risk score according to the algorithm does not mean that they are low or medium risk gamblers. It simply means that there is less certainty (i.e., detectable evidence) that such customers are problem gamblers. No performance measures were presented by the PwC/RGC analysts for the algorithm's ability to correctly place people into the various risk categories. Nor did the approach account for the fact that the majority of problem gamblers would not be detected by the model. Therefore, it was not appropriate to re-assign false negatives to lower risk categories.

This is a very important distinction; most high-risk and problem gamblers will not be detected by the models, and thus, are likely to fall into these lower categories. It is necessary to develop 'No Risk' models for detecting behaviour patterns that are associated specifically with no or low-risk outcomes. The players categorized as "Low Risk" by the algorithm could then be placed in the 'No Risk' category with confidence that very few of these players were "High Risk" or problem gamblers.

The analysts did not have the appropriate concept of risk. They need to generate risk profiles for those customers categorized as problem gamblers or no-risk gamblers to confirm most of those so designated fit that profile.

3.4 The assumption that there are a set of universal markers that can be used similarly across operators, game types and seasons to uniformly identify risk

Based on our research and experience to date at Focal Research, we have not found it possible to generate a universal model that performs similarly across products and operators. This was noted in our original paper published in 2011[5] and then systematically assessed as part of our collaborative research project initiated with NCF partners in 2013.

As the PwC/RGC authors reported multiple times throughout their report a “one-size-fits-all” approach is not advisable. This was cited as one of the reasons for adopting a customer-segmentation approach for customizing models. Unfortunately, they did not apply this same reasoning and testing to the markers, or, for that matter, to the different operators. If one-size does not fit all for different customer groups, this concept might also apply to operators. Because of differences in tracking systems, retained data sets, RG practices and product offerings there are distinct differences in terms of the markers that can be generated and the degree of model precision achieved for different operators.

Simply reducing risk detection to a limited set of common markers in the interest of achieving parity is not enough; the analysts needed to provide evidence that such markers achieved those objectives across operators. This analysis is missing from the PwC/RGC report. Presumably this was the reason for including multiple operators and indeed the reported results suggest that the analysts were aware that operator differences contributed to some limitations in the datasets. Operators that do have access to more data should not be penalized through the use of less efficient markers nor should differences be ignored in business practices or product offerings which pose lesser or greater risk to patrons. This is especially relevant if social responsibility, risk management and sustainable practices are to be positioned as competitive advantages for operators in growing and managing their business.

It should be noted that using standardized behavior inputs to uniformly identify high-intensity players (e.g., outliers) means that a similar percent of customers will be identified for interactions in all locations regardless of actual customer risk profiles for such operators. Such an approach will not be helpful in determining what policy, practices or actions actually assist gamblers in managing or mitigating risk and harm and which factors increase risk or have no impact. Essentially, a similar proportion of extreme behaviours will always be flagged for interaction by such a model if it is based on the use of standardized distributions.

---

[5] Schellinck, T., & Schrans, T. (2011). Intelligent design: How to model gambler risk assessment by using loyalty tracking data. Journal of Gambling Issues, pp. 51–68.

3.5 The Use of 'Daily Triggers' to detect in-play risk is promising but additional analysis is required before the value of the current markers can be assessed

The concept of developing daily triggers is an innovative and promising approach to identifying those at risk due to gambling. We recommend that such markers be created using relative values rather than absolute values if possible. From Tables 11.1-11.4 we see that the daily trigger variables are based on absolute amounts of bet value differences between betting days. It is recommended that these variables instead be based on relative amounts such as the average big win for a particular player followed by a percentage increase in the following bet day. This method avoids having the marker tied to specific amounts. Average bet values will be different for different operators, for gamblers in different countries or markets, and can be expected to change. Variables based on relative amounts should be more robust over time and across applications.

3.6 The concept of 'Graded Intervention' is reasonable but as described in the report will be impractical to implement or manage

The concept of graded intervention is a good one but it is not feasible as presented. The analysts illustrate the possible application of graded intervention (Fig. 39; p. 67) assuming the top 20% of players identified could be limited in how much they can bet. As this is 20% of an operator's active player base an operator that has 500,000 regular players could be expected to limit the play of 100,000 customers, the vast majority of whom will not be problem gamblers. While we realize the figures quoted in the report are supposed to be illustrative, the rationale provided does not make a good case for graded interventions when it presents such an unrealistic and impractical example.

Issue 4. Assessment of Conclusions

The PwC authors summarized their findings in the form of answering specific research questions (p. 65). Given our concerns regarding the analyses, we have reservations about the reliability and validity of their conclusions. We have repeated their questions below and provided comments based on our previous experience and review of the analysis carried out in the report. Among our concerns: the likelihood of a low rate of false positives; the degree of stability and reliability of the model across segments and operators; and the lack of a test sample to measure performance.

4.1 Can remote problem gamblers be identified by their online transactional behaviour?

We agree that problem gamblers can be identified online however, the lack of proper testing as described in Sections 1 through 3 make it difficult to evaluate whether the markers identified in the PwC/RGC study could be used effectively for this purpose.

4.2 How soon can operators identify remote problem gamblers in their customer lifecycle?

The analysts made the observation that the model stabilized after as little as three months. We caution that these results were largely due to the lack of complexity in the markers used in the models. When using intensity based markers it would be relatively easy to pick up the most frequent, regular customers during this window but again the degree of false positive identification associated with the exclusive use of such markers could make this an inefficient method for prioritizing customer action. It is unclear if the model would be flagging new customers each month or simply identifying the same customers falling in the high-turnover tiers over and over again.

4.3 Do markers of remote problem gambling vary for different groups of customers?

This research found some differences in markers by segment. What is not clear in this report is how stable and reliable the customer segments identified are. We do not know, for example, if the model applies to the target group consistently, is relevant for most operators and relatively stable across operators. We need this information to ensure that the appropriate algorithms are applied to appropriate customer groups. In the current research we have no idea about the stability and reliability of the customer segments produced.[6]

One important point to note is that there may be a strong seasonality effect in gambling with the nature of betting changing depending on the time of year. This possibility was not taken into account in this report. We have found that an 'all year' model is unlikely to be as effective as one that takes into account product seasonality, especially in relationship to sports betting activity.

4.4 Could operators identify a remote problem gambler 'in-the-moment'?

The daily triggers research may identify gamblers earlier than other models. Nonetheless, additional analysis is required to explore the issue more fully. Again, the existing dataset could be used to simulate outcomes to determine how many different customers would be identified

---

[6] Sarstedt, M. & Mooi, E. (2014). Cluster Analysis. In *A Concise Guide to Market Research: the process, data and methods using IBM SPSS* (2nd ed.) Statistics Berlin: Springer-Verlag. DOI 10.1007/9783642539640.

each day and what proportion are new or repeat identifications. This analysis was not undertaken and/or reported on in the PwC/RGC report and the reported findings were restricted to examples of a few specific customers only.

4.5 What markers are practical to implement online, especially given the level of false positives for those predicted as remote problem gamblers?

Although the authors point out their concern regarding false positive rates (p. 66), their methodology has likely lead to false positive estimates that are far lower than is actually the case. This is of serious concern. They did not use a test sample to produce performance metrics; they excluded the low and medium risk gamblers when producing the performance metrics; they reported accuracy for one period rather than for the whole year.

Issue 5. Assessment of Report Recommendations and Limitations

Based upon their work, the authors made eight recommendations (p. 68) that they believe should be followed by operators in their efforts to minimise gambling harm. They also addressed the limitations of their work. As they plan to apply these principles in their future model development, we have addressed each of these in turn and provided our assessment of the strengths and weaknesses of their approach.

5.1 Use self-reported PGSI survey results to train detection models, not self-exclusion data.

This is an excellent first point to make. The evidence from Phase I and the analysis of this survey data strongly supports this conclusion as has our experience in using the PGSI and FLAGS (FocaL Adult Gambling Screen) for the past 15 years. [7]

[7] Focal Research has published a series of peer reviewed papers associated with the development of a validated instrument for detecting pre-harm risk, as well as, harm and problem gambling including: Schellinck, T., Schrans, T., & Collins, P. (2007) The Concept of Formative Constructs and Suggested Applications in the Area of Gambling Risk Assessment: An Annotated Bibliography, Ontario Problem Gambling Research Centre (OPGRC); Schellinck, T. (2006) Phase II: Quantitative Analysis for the Victoria Self-Administered Problem Gambling Screen, Victoria Department of Justice; Schellinck, T., Schrans, T., Bliemel, M., Schellinck, H., & Ontario Ministry of Health and Long-Term Care (2011b). Raising The FLAGS: A Pilot Study Adapting FLAGS, A Next Generation Gambling Risk Assessment Instrument, For Use in Identifying Risk Among General Gambling; Schellinck, T., Schrans, T., Bliemel, M., & Schellinck, H. M. (2015a). Construct Development for the FocaL Adult Gambling Screen for Electronic Gambling Machine players (FLAGS-EGM): A Measurement Instrument for Risk due to Gambling Harm and Problem Gambling Associated with Electronic Gambling Machines. Journal of Gambling Issues, 140–173; Schellinck, T., Schrans, T., Bliemel, M., & Schellinck, H. M (May 2015) Instrument Development for the FocaL Adult Gambling Screen (FLAGS-EGM): A Measurement of Risk and Problem Gambling Associated with Electronic Gambling Machines., Journal of Gambling Issues, 30, 174-200

5.2 Use a range of data sets in detection models, including demographics (e.g., age, gender), account activity (e.g., deposits, withdrawals, use of protection tools), play activity (e.g. volume of bets), and customer service contacts.

We agree that a variety of algorithms using different sets of variables, when combined can lead to improved at-risk gambler detection. In this report, the account activity variables proved predictive. When we applied similar variables to a comparable dataset, we found that all of those identified by the algorithm *were in the operator's top 80 - 100 % of turnover customers.* The current set of markers used by analysts contain too many intensity variables or markers associated with high turnover, and therefore those problem gamblers in the bottom 80% turnover segments are not likely to be identified. In other words, the algorithms developed in the report are likely to target the operator's VIP customers and ignore risk among the majority of customers. Table 8.1 in Appendix 8 shows that the three most predictive variables are bet volume, bet value and deposit frequency, all measures of gambling intensity rather than behaviour pattern variables. This is why the models identify only those in the very top turnover tier.

The available data varies across operators. The best models use all available information/data that are proven to be useful in identifying at-risk gamblers not simply a set of common markers that can be easily generated across all operators as was done in this report. Such an approach may offer little value in reflecting the impact of operator practices and policy in mitigating risk.

5.3 Use customer segmentation based on play behaviour to identify higher risk gambling groups

The creation of segments and then developing segment specific algorithms has merit. It is, however, unclear how representative the segments are of the players at various sites. The player profile can vary greatly among operators so the segments may be useful for some but may be ineffective for others. As well, the offerings of online operators are expected to change over time so the veracity of the segments can be expected to change as well. It is advantageous to have serial algorithms that narrow down the process of distinguishing at-risk gamblers from those who are low risk gamblers, but the effectiveness of the algorithms at each step must be proven stable. The authors' first step was to segment the market into customer grouping and then develop the algorithms to further separate the at-risk gamblers. The basis for this first clustering step was not established nor was it proven to be robust over time and across operators.

5.4 Use multivariate models to capture complex combinations of features

The authors point out that they were not trying to develop a final algorithm but to demonstrate that one could be designed for online operators. In this they were successful although the reported performance of these algorithms is greatly exaggerated and has limited application.

5.5 Run detection models from the day the account is created, starting with demographic data to identify higher risk groups.

We do not yet know the performance metrics of the demographic data algorithms though they may have promise. It is possible that the false positive rate will be so high that it will not be reasonable for most operators to use such a detection model from 'day one'. It is too early to recommend that detection models run from the day the account is created. There are also issues associated with customer profiling and algorithmic bias that must be considered when using demographic, socio-economic and or other ethnicity indicators.

5.6 Update detection models daily to create customer risk scores that change as play behaviour develops over time.

This approach may work but may be no better than a system that runs monthly and it will be a lot more expensive to operate. Most of those players identified daily will be the same people who were identified on previous days and it makes sense to accumulate confirmatory 'hits' before action is taken; otherwise, the impact on operator resources could be considerable, distracting them from more meaningful customer action and support. The long-term data on the player will be available in any case, including past output of the algorithms in categorizing the player as at-risk or not. In short, the daily information is likely to be used in conjunction with the longer-term data. The authors have the data in hand to explore some of these issues and make a more informed argument for or against continuous, daily or monthly runs of the algorithms.

5.7 Complement this with daily detection of specific problem gambling triggers to enable immediate investigation and potentially intervention.

The daily triggers, as designed and tested presently, would only be triggered at the end of a day's activity. A system that scores the person in advance of the day's activity can be used to reach out to contact the person during or after a moment of crisis. The daily detection system may have merit if it identifies people who would never be identified using the other algorithms. However, those variables that trigger using two days of play (such as chasing large wins) can also be incorporated into the other algorithms and used to identify at-risk gamblers at any interval felt appropriate by the operator.

As noted earlier**,** the variables should not rely on absolute value thresholds, i.e., pounds bet, but should instead be relative amounts such as percent increase in spending the next betting day.

5.8 Use risk thresholds customised by segment to set detection and intervention policies.

We agree that the operators should examine the use of different thresholds for different forms of interaction with the customers. If they are using personal contact then any threshold above 1% - 5% will likely be impractical for an operator in terms of resources. The authors could have tested and reported on thresholds at various levels to explore impacts. For example, a reduced threshold would drastically reduce the hit-rates but increase the precision and stakeholders need to see what is possible using these algorithms to find at-risk gamblers. We feel the response will be that the algorithms are useful, but they are certainly not the silver bullet that solves the problem of finding and helping all at-risk gamblers.

5.9 Limitations

The following section addresses the limitations noted by the authors of the PwC/RGC report. The authors correctly cite unknown biases due to low response rate. However, we feel the 6.5% response rate is much lower than necessary.[8]

It is noted as a limitation that operators will have different sets of data (i.e., the number and definition of the variables supplied) that they can provide for algorithm development. Our experience is that those operators who supply fewer variables, or whose data is only in summary form, will likely have less effective algorithms. This suggests that at-risk gambler detection systems based on algorithms need to be designed/customized for each operator if they intend to create the most accurate detection system possible. It is in the interest of the operators to do this as less accurate algorithms are less efficient; they will have to contact more people to find at-risk gamblers and they will also be contacting more people who are not at-risk, not to mention missing risk among certain player segments.

The authors cite the fact that they could not measure the accuracy of the algorithms in identifying problem gamblers who play on different sites as a limitation. We feel they have the data in hand to analyze this issue (as discussed previously) to see if it is a real limitation.

[8] Focal Research regularly achieves response rates for online surveys with gambling customers ranging from 12% to 40% using similar survey techniques suggesting there may be methods for supporting survey completion rates.

The report states "*Furthermore, given the inherent bias during sampling towards active customers, many of the conclusions for the present study cannot be generalised to the larger population of online gamblers at this stage.*" In fact, it should not be generalized to the larger population of online gamblers. When the system of algorithms is implemented on an operator's site it should only be applied to those customers who are active gamblers, using the same definition as was used to select the sample of 160,000 players eligible to be surveyed. Those at-risk gamblers who are low activity players will be missed but they make up a very small proportion of all at-risk gamblers; moreover, they are unlikely to have sufficient play records to reliably assess the risk of such behavior.

5.10 Recommendations for Phase 3

The authors indicate they have not developed final models, and in this review, we have described many opportunities for improvements to analysis, marker design and segmentation. Much more work is required before final algorithms can be created. The report states that "The overall objective of a Phase 3 is to develop, test and refine an intervention strategy for minimising harm from remote gambling in the UK" (P. 70). In our view, it is premature to recommend using these algorithms in Phase 3 of the research, and would likely lead to incorrect conclusions. We feel there is room for substantial improvement to the algorithms and that they need to be tested *in situ* before they are used as a means for testing and informing intervention strategies.

Focal Research's Recommendation for Next Steps

The PwC/RGC study has produced a rich dataset that, with additional analysis, should yield valuable insight into the efficacy of using online gambling data to identify at-risk remote gamblers. There were many insightful analyses conducted by the PwC/RGC team that are creative and potentially add knowledge of how the algorithms could be developed and applied to the task. However, as discussed, there are also shortfalls to be addressed that have implications for the practical value of the results in informing ongoing work in this area.

The analysts used a two-step process to derive the algorithms. The first algorithm was developed using a combined data set and unsupervised clustering to create nine player segments, four of which eventually contained 78% of problem gamblers retained in the analysis. Only those players who were members of these four 'clusters' were then used to create the final algorithms. The analysts used two split-halves of the data to test the stability of the results and the consistency of the variables used to create the clusters. However, there is no indication the analysts tested the stability or relevance of the initial clustering algorithm by applying it to samples for each of the individual operators taking part in the study. Nor was the performance of the final algorithms assessed using the total overall combined player samples or even the full cluster samples (e.g., including all players not just those scoring PGSI=0 or 8+ in each cluster). Consequently, two important characteristics of the clusters that can be expected to vary were not tested in the current study; the proportion of problem gamblers falling in each cluster and the size of each of the clusters within an operator's customer base.

When applied to a representative player population, it is possible, for example, that only 60% of problem gamblers will fall into any of the four designated clusters as compared to the 78% in the non-representative survey sample used to determine the clusters. The size of the four clusters as a proportion of the full target customer base will also likely vary by operator and possibly over time. If the pre-selected clusters account for a smaller percentage of the total customer base for a particular operator, then this too will reduce the number of problem gamblers that may fall into any one of these pre-selected clusters. Consequently, the first phase clustering algorithm will be ineffective in isolating problem gamblers and the final algorithms developed based on the original clusters will identify fewer problem gamblers. As a result the overall effectiveness of the proposed approach is substantially overstated in the report.

It is important to understand that the nine segment solution presented is only one of several possible solutions. It is unclear from the report how many cluster analyses were explored by the PwC analysts or how the solution presented was selected but it is reasonable to assume that the clusters and the criteria for assigning members to those final clusters were adjusted until an optimal solution was found. If this is the case then this particular clustering solution has limited application and is specific to this particular sample of players. When applied to another sample it will almost invariably provide less power in clustering problem gamblers as described above, and those problem gamblers it does identify may have a different profile rendering subsequent algorithms less effective than reported.

Accurate algorithms when embedded in a strong RG program can be an effective tool to assist operators in targeting and managing resources for player protection and prevention. At best the current research has identified a limited potential set of variables that were found to be effective in detecting some problem gamblers as compared to 'no-problem' gamblers within certain player sub-segments (e.g., high frequency-high intensity gamblers).

The issues identified with sampling, limited marker or variable coverage, analysis and interpretation raise questions about the wider practical value of the reported results and conclusions. Additional analysis of this dataset could address many of the issues raised and is encouraged to help inform ongoing work in this growing area of gambling research.

A more detailed review of the PwC report may be found at FocalResearch.com.