A comparison of attribute-focused and harm-focused methods for assessing the risk of organised crime groups: Are they in agreement?

Paul Mulholland

Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA. Email: paul.mulholland@open.ac.uk

Terri Cole

Department of Psychology, University of Bournemouth, Poole House P116, Talbot Campus, Fern Barrow, Poole, BH12 5BB. Email: tcole@bournemouth.ac.uk

Abstract Risk assessment methods are used to map the threat posed by organised crime and assist in the prioritisation of policing resources. Within policing, risk assessment methods are currently in use that either focus on the attributes of organised crime groups or the types of harm posed. Statistical analysis was undertaken of risk assessment data produced using one attribute-focused and one harm-focused method. The study was based on a non-random sample of 113 attribute-focused and 77 harm-focused risk assessments of the same 17 OCGs, and a sub-sample of 36 pairs of risk assessments of 16 of the OCGs. A moderate correlation was found between the two methods, despite differences in their risk assessment processes. Lower levels of agreement tended to be associated OCGs posing high levels of harm, particularly financial and psychological. Using the two risk assessment methods in combination is proposed as a way of building consensus that draws on their alternative perspectives. The techniques applied in this analysis could be used to inform revisions to the methods and their supporting materials, identify outlying risk assessments, and potentially incorporate other information sources such as harm indices.

Introduction

Organised crime can be defined as ‘a continuing criminal enterprise that rationally works to profit from illicit activities that are often in great public demand. Its continuing existence is maintained through the use of force, threats, monopoly control and/or the corruption of public officials’ (Albanese, 2008, p263). The traditional view that Organised Crime Groups (OCGs) engaged in such activities always operate as a hierarchical, stable, “family business” no longer holds. Organised Crime may be ‘undertaken by professional and non-professional criminals who come together in fluid entrepreneurial networks which in turn are characterised by various often ad hoc and opportunistic types of organised structure’ (Hamilton-Smith and Mackenzie, 2010, p264). OCGs are also increasingly involved in multiple types of criminality. Europol (2017) estimate the proportion of OCGs engaged in more than one criminal activity has increased from 33% in 2013 to 45% in 2017.

New technology is changing the nature of organised crime and criminality more generally. Anderson et al (2013) distinguish between traditional crimes that are now conducted online (e.g. tax fraud), transitional crimes whose modus operandi has changed in the move to online (e.g. credit card fraud) and new forms of crime made possible by new technology (e.g. denial-of-service attacks). Over time there have also been changes in the nature of criminality and how they are defined in legalisation, for example, the UK’s Modern Slavery Act 2015.

The risk assessment of organised crime can be carried out using different methods and for different purposes. Risk assessment can be conducted to understand the OCG threat related to a particular crime type (e.g. drug trafficking) or associated with a particular geographical area. These forms of assessment are generally referred to as Organised Crime Threat Assessments (Zoutendijk, 2010). Risk assessment can also be carried out on the level of individual OCGs. This form of assessment can be used, for example, to assist in the prioritisation of policing resources as well as contributing to a mapping of the overall OCG threat.

Methods for assessing the risk posed by individual OCGs often characterise OCGs in terms of their risk, harm and threat (Hamilton-Smith and Mackenzie, 2010). Harm is a measure of the possible negative impact of the OCG. Threat is a measure of the probability of the harm being realised. Risk is calculated as a combination of harm and threat. The assessments of threat and, in particular harm, from which the risk assessments are derived, can be carried out in two main ways. Hamilton-Smith and Mackenzie (2010) distinguish these as attribute-focused and harm-focused risk assessment.

Attribute-focused risk assessment

Attribute-focused risk assessment methods describe an OCG in terms of its attributes, for example, the types of criminality in which the OCG is engaged (e.g. drugs, human trafficking), the scale of involvement (e.g. frequency, geographical scope) and the capabilities of the OCG (e.g. access to weapons, involvement of corrupt officials). The Sleipnir method developed by the Royal Canadian Mounted Police (RCMP), described by Beken (2004) and Hamilton-Smith and Mackenzie (2010), assigns a level (low, medium or high) to 19 different OCG attributes (e.g. corruption, violence, infiltration, expertise). Klerks, described by Beken (2004) and Hamilton-Smith and Mackenzie (2010), uses scores on 31 attributes in a complex calculation to determine the risk level of an OCG. Organised Crime Group Mapping (OCGM), discussed by Crocker et al (2016) and Carr et al (2017), provides a framework for mapping or indexing OCGs in terms of their intent and capability to cause harm.

Hamilton-Smith and Mackenzie (2010) point out that attribute-focused risk assessments such as Sleipnir can lack an empirical basis for the weighting applied to the scores for different crime types of varying involvement or severity. They also point out that the number of attributes against which the OCG may have to be described can make the process time consuming. The method by which risk is calculated from the scores for crime types can also be complex and opaque.

Harm-focused risk assessment

Harm-focused assessment methods describe the risk posed by an OCG using higher level thematic harm categories rather than the specific types of criminality in which the OCG is engaged. The Metropolitan Police Service (MPS) Criminal Networks (CN) Prioritisation Matrix, discussed by Hamilton-Smith and Mackenzie (2010), distinguished three main categories of harm: social, economic and political. Using the Prioritisation Matrix, an OCG is initially scored in terms of its level of involvement (low, medium, high) for each of a number of crime types. Each crime type is associated with a primary and possibly secondary thematic harm type. For example, vehicle crime is primarily associated with economic harm and secondarily associated with social harm. These scores are then combined to produce a harm level for each of the three harm categories. Management of Risk in Law Enforcement (MoRiLE) (Dowden, 2017) can be used to calculate the harm of an OCG according to a number of factors (e.g. physical, psychological, financial) which are added to produce an overall harm score. The harm score is combined with a likelihood score to produce a risk score for the OCG.

As harm-focused methods of risk assessment use higher level thematic categories and a particular type of criminality may contribute to multiple types of harm to a varying extent, there can be ambiguities and uncertainties in the assessment process, as well as the potential for double accounting of criminalities across multiple harm types (Hamilton-Smith and Mackenzie, 2010).

Aim of the study

Attribute-focused and harm-focused methods of risk assessment for OCGs are both in active use in policing. The aim of this study was to identify the level of agreement between risk assessments made using the two methods. Prior to this study, there had never been a comparison of risk assessments of the same OCGs using both an attribute-focused and harm-focused risk assessment method.

Comparison of two risk assessment methods

This study compared OCG risk assessments conducted using one attribute-focused and one harm-focused risk assessment method. The methods compared were the attributed-focused OCGM and the harm-focused MoRiLE. OCGM and MoRiLE differ both in terms of the process by which a risk assessment is derived and the terminology used to describe the process.

The OCGM assessment process involves describing the OCG according to a series of attributes. These attributes are organised into two groups. One group of attributes describe the criminality of the OCG. The other group describes its intent and capability. For criminality, the OCG is given a score for each type of criminal activity related to its operation, for example, drugs, fraud, sexual offences and environmental crime. For each type of criminality in which the OCG is involved, the level of activity is graded as low, medium or high. These grades are translated into a numerical value. The values associated with low, medium and high levels of involvement in different levels of criminal activity are weighted. The weights are determined by drawing on available sources of information related to the measurement of harm. These include estimating harm from the average sentence given for a particular offence (Babyak et al, 2009; Bangs, 2016; Sullivan and Su-Wuen 2012), the use of sentencing guidelines as a measure of the harm caused by a crime (Sherman et al, 2016; Ratcliffe, 2015) and estimates of the financial costs of crime (Dubourg and Prichard, 2007; Levi, 2016; Anderson et al, 2013). The weighted scores for each type of criminal activity are summed to produce an overall criminality score for the OCG. The OCG is also described according to a number of attributes related to its intent and capability, for example, growth, resources and expertise. Each relevant intent and capability attribute is scored on a weighted three-point scale. The sum of these scores produces an overall intent and capability score for the OCG. The OCGM score for the OCG is the sum of the scores for criminality and intent and capability. The OCGM score and sub-scores for criminality and intent and capability can be classified into bands depending on their scale.

The MoRiLE assessment process involves describing an OCG in terms of a set of thematic harm types: physical, psychological, financial, community, environmental and geographic scope. The last of the themes therefore relates to the scope of the harm rather than the type of harm caused. For each theme the OCG is given a score on a weighted seven-point scale ranging from none/negligible to catastrophic. The scores for each harm type are summed to produce an overall harm score for the OCG. A likelihood score is calculated for the OCG using a set of thematic factors related to the intent and capability of the OCG, the frequency and volume of criminality and potential victim vulnerability. Each factor is scored on a weighted five-point scale. The risk score is calculated as the mathematical product of the harm and likelihood scores. MoRiLE can also be used to specify an operational score with respect to the OCG. This considers, for example, the required capacity and capability to address the risk posed by the OCG. The operational score is used in combination with the risk score to prioritise (i.e. grade) the OCG.

Hamilton-Smith and Mackenzie (2010) in their review of organised crime assessment methods, draw on Brown's (1988) model of risk assessment (see figure 1). According to the model, the risk of an OCG is a function of harm and threat. Harm is a measure of the possible negative impact of the OCG. Threat is a measure of the probability of the harm being realised. Threat itself can be measured as a function of both the intent and capability of the OCG.

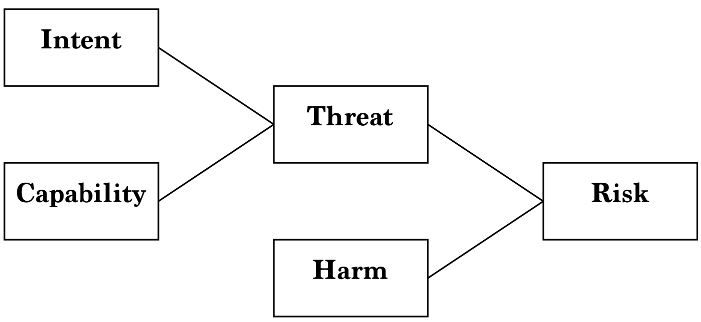

Figure 1: The relationship between risk, threat, harm, intent and capability as proposed by Brown (1988).

Brown's (1988) model can be used to align the terminology of OCGM and MoRiLE (see table 1). The OCGM score equates with the risk score of MoRiLE. In OCGM, harm is characterised as criminality and threat is characterised as intent and capability. In MoRiLE, threat is described as likelihood. The operational score and grading produced by MoRiLE is beyond this model of risk, harm and threat and is not covered in OCGM. For clarity, and to assist comparison of the two methods, we will adopt the terminology of Brown's (1988) model rather than OCGM and MoRiLE specific terminology, for example, referring to OCGM threat and MoRiLE threat rather than OCGM intent and capability and MoRiLE likelihood.

Table 1: A comparison of OCGM and MoRiLE terminology.

	OCGM	**MoRiLE**
Risk	OCGM score and band that combines scores for criminality, and intent and capability.	Risk score calculated from scores for harm and likelihood.
Harm	Criminality score and band calculated from attributes.	Harm score calculated from scores for harm types.
Threat	Intent and capability score and band calculated from attributes.	Likelihood score calculated from threat, scale and victim vulnerability factors.
Operational position	Outside the scope of OCGM.	Operational score calculated from policing capability and capacity factors. Grading of the OCG calculated from the operational score and risk score.

Analysis of the risk assessment dataset

The aim of this research was to compare the level of agreement between attribute-focused and harm-focused methods of risk assessment, specifically OCGM and MoRiLE. To enable this analysis, the policing partner within the research project, sourced OCGM and MoRiLE risk assessments that had been carried out on the same OCGs. This search produced an initial dataset of OCGM and MoRiLE risk assessments (see table 2). The OCGM data comprised 113 risk assessments related to 17 OCGs. The MoRiLE data comprised 77 risk assessments related to the same 17 OCGs. The MoRiLE data included the full scoring for each assessment. The OCGM data included the overall risk, harm and threat scores. It did not include a further breakdown of harm and threat attributes. Information was not provided on who had conducted the OCGM or MoRiLE assessments or the police intelligence on which those assessments were based. The MoRiLE assessments were all carried out within an 11-month period. The OCGM assessments were carried out over a 6-year period. 65 of the OCGM assessments overlapped with the 11-month period of the MoRiLE assessments.

Table 2: Descriptive statistics for the initial dataset.

Risk assessment method	**Mean risk score**	**N**	**SD**	**Minimum**	**Maximum**
OCGM	57.33	113	44.04	0	198
MoRiLE	91.7	77	96.58	0	504

For the OCGM data, the risk score ranged from 0 to 198 with a mean score of 57.3. OCGM scores varied both within and across OCGs (see figure 2). The mean scores for individual OCGs ranged from 15 to 132.

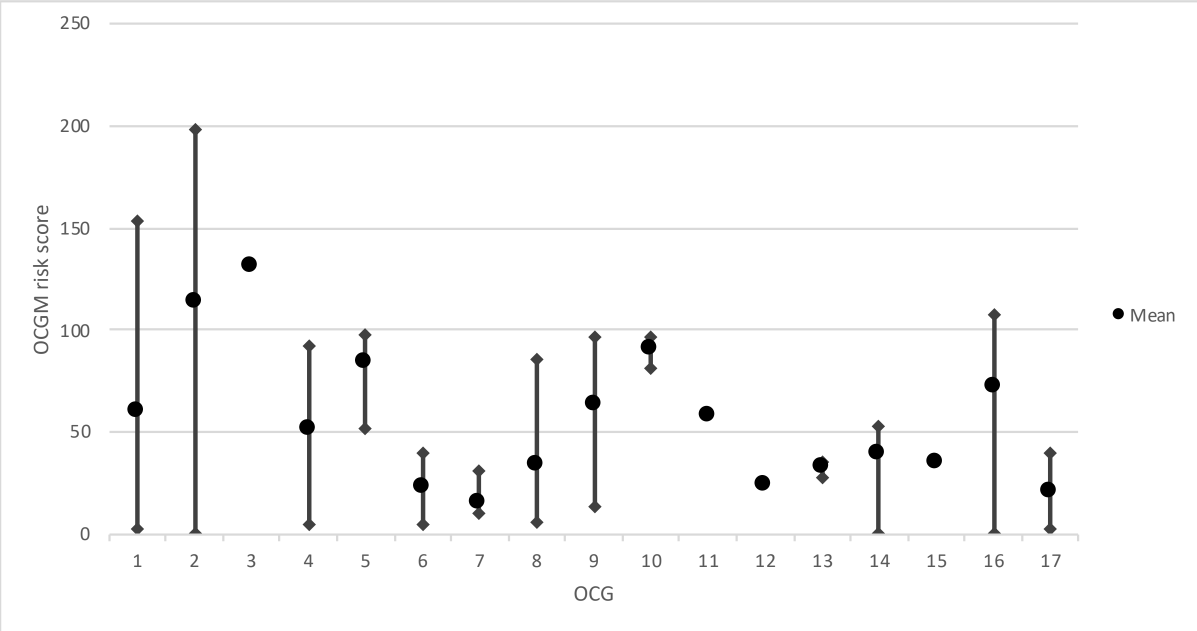

Figure 2: OCGM risk scores. For each OCG (shown along the x-axis) the line connects the minimum and maximum scores. The dot indicates the mean score.

MoRiLE risk scores for the OCGs ranged from 0 to 504 with a mean score of 91.7. MoRiLE scores varied both within and across OCGs (see figure 3). The variation across the set of MoRiLE scores (N = 77) was greater than that for the OCGM scores (N = 133). The Coefficient of Variation (a measure of variation calculated as the standard deviation divided by the mean) was larger for MoRiLE (1.05) than OCGM (0.77).

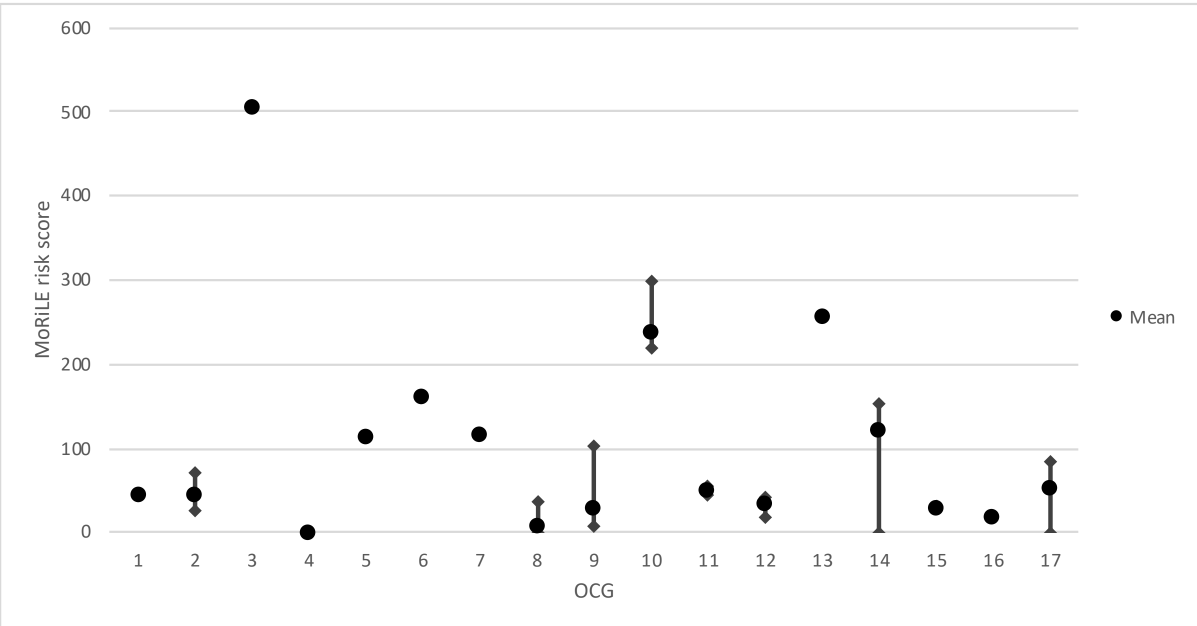

Figure 3: MoRiLE risk scores. For each OCG (shown along the x-axis) the line connects the minimum and maximum scores. The dot indicates the mean score.

The MoRiLE thematic harm scores for the 17 OCGs give some indication as to the nature of criminal activity in which the OCGs were engaged. On average the OCGs were assessed as causing some level of harm in 3.5 of the 6 thematic harm types. This may indicate involvement in multiple criminal activities and/or involvement in a criminal activity that causes multiple types of harm. 88% of the OCGs had been assessed as operating at some broader level of geographical scope. Varying proportions of the OCGs had involvement in criminal activities causing physical (47%), psychological (59%), financial (71%), community (76%) and environmental (12%) harm.

Pairing the OCGM and MoRiLE assessments

It can be seen from the above analysis that the risk scores (whether calculated by OCGM and MoRiLE) can vary widely across individual assessments. This means that when looking to compare OCGM and MoRiLE, we cannot just compare across the available assessments for each OCG as the risk level could have changed considerably between individual MoRiLE and OCGM assessments. For this reason, a method was adopted that paired OCGM and MoRiLE assessments of an OCG that were reasonably close in time.

From the initial dataset, a smaller dataset was selected in which OCGM and MoRiLE assessments were matched into pairs. OCGM and MoRiLE assessments were matched into pairs if:

- They concerned the same OCG;
- Chronologically they were the closest possible match between the available MoRiLE and OCGM assessments of the OCG (i.e. the OCGM assessment was not closer in time to another MoRiLE assessment and vice versa);
- The timespan between the OCGM and MoRiLE assessment dates was less than four months.

Four months was chosen as a cut-off heuristically to keep the assessments relatively close in time while not overly reducing the size of the dataset available for analysis. To test this heuristic, later analysis will investigate whether the time difference between the two assessments affects the relationship between the two scores.

Description of the paired dataset

The selection process produced 36 pairs of OCGM and MoRiLE scores. The 36 pairs covered 16 different OCGs. Between one and four pairs were associated with each OCG. The selected dataset therefore comprises 31.9% (36 out of 113) of the OCGM scores and 46.8% (36 out of 77) of the MoRiLE scores in the initial dataset.

The mean difference between the two assessment dates of a pair was 30.3 days. The minimum was 0 days (i.e. OCGM and MoRiLE assessments done on the same day) and the maximum was 115 days. In the 36 pairs, the OCGM overall risk scores had a mean of 43.97 and varied from 0 to 154. MoRiLE risk scores had a mean of 89.5 and varied from 0 to 504. In comparing the descriptive statistics, the paired dataset (table 3) is similar to, and can therefore be considered representative of, the larger dataset (table 2).

Table 3: Descriptive statistics for the paired dataset.

Risk assessment method	Mean risk score	N	SD	Minimum	Maximum
OCGM	43.97	36	44.22	0	154
MoRiLE	89.5	36	110.45	0	504

The distribution of OCGM risk scores is shown in figure 4. A similar distribution pattern is found for the MoRiLE scores (figure 5), both having a high frequency of lower risk scores.

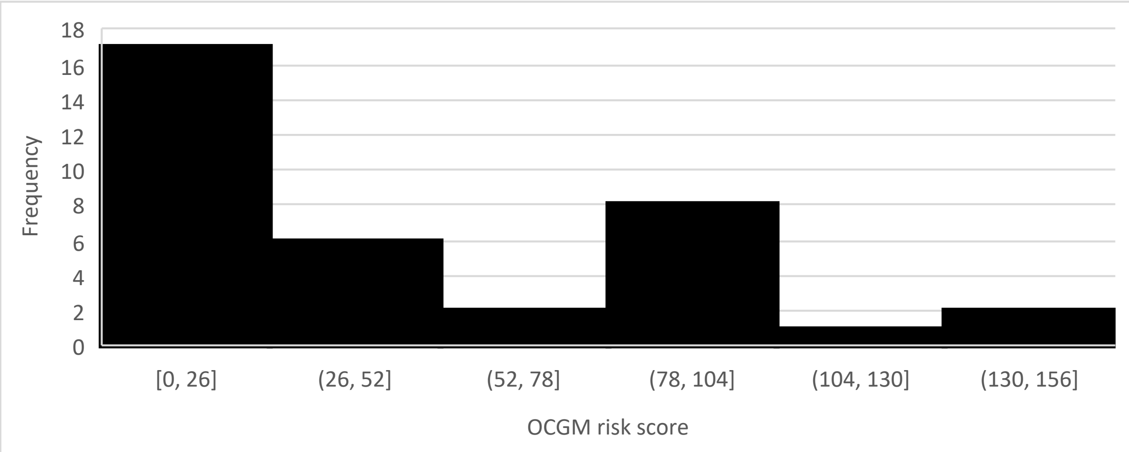

Figure 4: Frequency distribution of OCGM risk scores in the paired dataset.

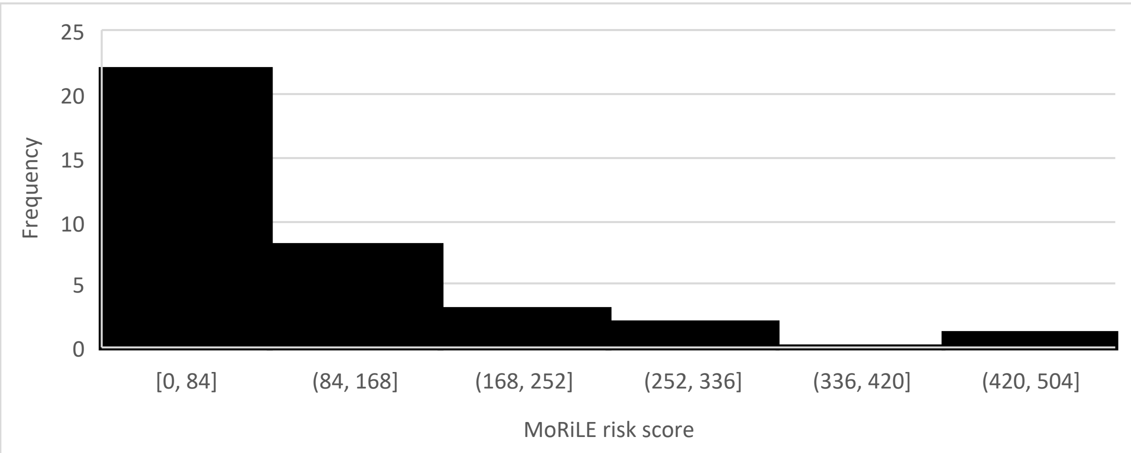

Figure 5: Frequency distribution of MoRiLE risk scores in the paired dataset.

Correlations for risk, harm and threat

Risk: A Shapiro-Wilk test confirmed that the OCGM risk scores ($p < 0.01$) and the MoRiLE risk scores ($p < 0.01$) were not normally distributed. As a consequence, the non-parametric Spearman correlation was used. A correlation was found between the OCGM and MoRiLE risk scores, $r(34) = 0.43$, $p < 0.01$. The correlation suggests that although OCGM and MoRiLE use different approaches there is a moderate correlation between scores produced by the two types of assessment.

A scatter plot of OCGM versus MoRiLE risk scores is shown in figure 6. The dotted line represents the linear trendline between the two sets of scores. The associated equation expresses the relationship between the MoRiLE risk score (x) and the predicted OCGM risk score (y) as visualised by the trendline. For each pair of scores, the difference was calculated between the actual OCGM score and the score predicted by the equation. The standard deviation of the differences between the actual and predicted OCGM scores was calculated as 39.46. Score pairs that are within one standard deviation of the trend line are shown as a back circle. Those between 1 and 1.5 standard deviations are show as a hollow circle. Those between 1.5 and 2 standard deviations are shown as a black triangle. Score pairs beyond 2 standard deviations are shown as a hollow triangle.

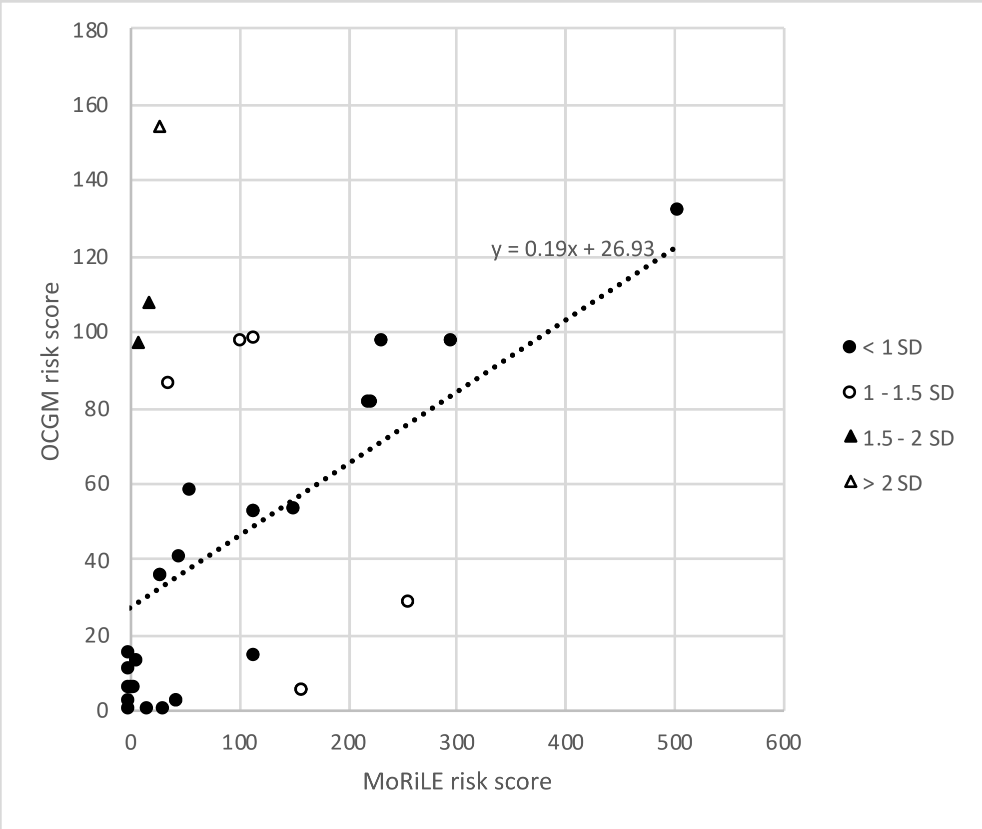

Figure 6: Comparison of OCGM and MoRiLE risk scores.

Harm: A Shapiro-Wilk test confirmed that the OCGM harm scores ($p < 0.01$) and MoRiLE harm scores ($p < 0.01$) were not normally distributed. A Spearman correlation was found between the OCGM and MoRiLE harm scores, $r(34) = 0.41$, $p < 0.05$. This suggests a moderate correlation between the OCGM and MoRiLE harm scores. A scatter plot of OCGM versus MoRiLE harm scores is shown in figure 7. The standard deviation of the difference between actual and predicted OCGM scores was calculated as 25.63. The chart illustrates two outliers (shown as hollow triangles) that are more than two standard deviations from the trendline. These represent the two highest OCGM harm scores and relatively low MoRiLE harm scores.

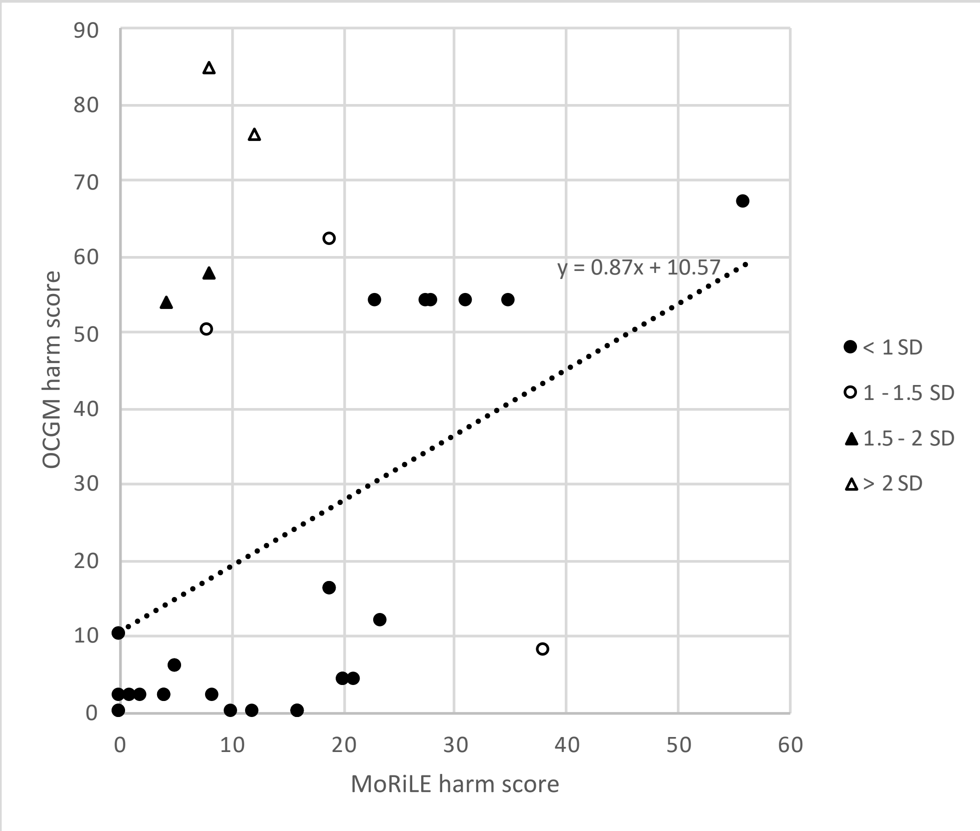

Figure 7: Comparison of OCGM and MoRiLE harm scores.

Threat: A Shapiro-Wilk test confirmed that the OCGM threat scores ($p < 0.01$) and MoRiLE threat scores ($p < 0.01$) were not normally distributed. A Spearman correlation was found between the OCGM and MoRiLE threat scores, $r(34) = 0.40$, $p < 0.05$. This suggests a moderate correlation between the OCGM and MoRiLE threat scores. A scatter plot of OCGM versus MoRiLE threat scores is shown in figure 8. The standard deviation of the difference between actual and predicted OCGM scores was calculated as 17.7. One datapoint (shown as a hollow triangle) is more than two standard deviations from the trendline.

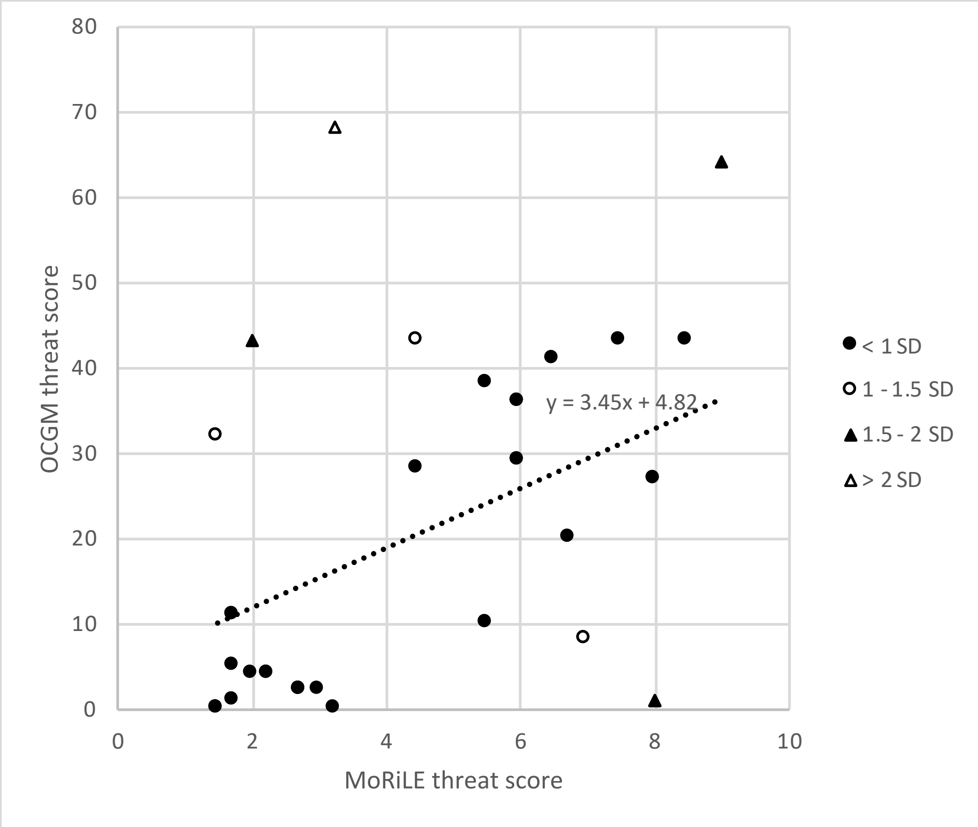

Figure 8: Comparison of OCGM and MoRiLE threat scores.

Effect of time difference between the two assessments

A test was conducted to determine whether the length of time between pairs of risk scores could be a factor in lowering the correlation between the scores. For each pair of scores, the absolute difference between the two scores and the number of days between the two scores was calculated. No significant correlation was found between the two measures, $r(34) = 0.19$, $p = 0.264$. However, any potential correlation between time difference and score difference could be reduced by the presence of stable OCGs whose risk scores do not change over time. To investigate this, three OCGs were removed from the correlation that had identical risk scores on temporally adjacent risk assessments. Even with the assessments related to these more stable OCGs removed, no correlation was found between the absolute difference in scores and the number of days between scores, $r(27) = 0.18$, $p = 0.356$. This suggests that within the selected dataset, in which pairs of scores are no more than four months apart, there is no significant effect due to the time difference between the two scores.

Analysis of outliers

A moderate correlation was found between the OCGM and MoRiLE assessments of risk, harm and threat. The two approaches could therefore lead in some cases to different prioritization decisions. For example, the highest OCGM risk score is paired with a relatively low MoRiLE risk score, which could justify different courses of police prioritization and action. Further analysis was conducted to identify the subset of paired assessments that would need to be excluded in order produce a strong correlation between the OCGM and MoRiLE scores. This subset of assessments was then further analysed in order to help identify reasons why the correlation between OCGM and MoRiLE was not stronger. The assessment pairs within this subset are referred to as outliers. It should be noted that these are not outliers in the sense of measurement errors to be removed prior to analysis. In this case the outliers have been included in the full analysis and have been subjected to further analysis into why the correlation is not stronger.

Whether a correlation can be considered to be sufficiently strong is dependent on context. When measuring agreement between human judgements (for example when determining inter-rater reliability), often a correlation of 0.7 or higher is taken to indicate a sufficient level of agreement. Therefore, an analysis was performed in which the risk, harm and threat correlations were repeated with assessment pairs removed that were more than 2, 1.5 and 1 standard deviation from the trendline (these are labelled on figures 6, 7 and 8). As the data in the correlation is restricted to assessment pairs increasingly near the trendline, the resulting correlation can be expected to increase in strength.

Spearman correlation tests with datapoints excluded that were more than 1, 1.5 and 2 standard deviations from the trendline are reported in table 4. It can be seen that, as expected, the strength of the correlations (as reported in the r values) increases as more assessment pairs are excluded. To reach a correlation of 0.7 for risk and threat, assessment pairs more than 1 standard deviation from the trendline need to be excluded. For each correlation in table 4, the Coefficient of Determination (R^2) is reported, which is calculated as the square of the r value. R^2 indicates the proportion of variance in the OCGM score that can be predicted from the MoRiLE score. An r value of greater than 0.7 (taken to indicate a strong correlation) is therefore also the point at which most of the variation in the OCGM score can be predicted from the MoRiLE score. In the complete paired dataset, 18% of the variance (i.e. $R^2= 0.18$) in the OCGM risk score can be predicted from the MoRiLE risk score. By contrast, with assessment pairs more than one standard deviation from the trendline excluded, 53% of the variance (i.e. $R^2= 0.53$) in the OCGM risk score can be predicted from the MoRiLE risk score.

Table 4: Correlation coefficients for risk, harm and threat excluding datapoints 2, 1.5 and 1 standard deviations from the trendline.

	Risk	**Harm**	**Threat**
Paired dataset	$r(34) = 0.43$, $p = 0.009$, with a $R^2 = 0.18$	$r(34) = 0.41$, $p = 0.013$, with a $R^2 = 0.17$	$r(34) = 0.40$, $p = 0.015$, with a $R^2 = 0.16$
> 2 SD excluded	$r(33) = 0.48$, $p = 0.004$, with a $R^2 = 0.23$	$r(32) = 0.49$, $p = 0.004$, with a $R^2 = 0.24$	$r(33) = 0.42$, $p = 0.011$, with a $R^2 = 0.18$
> 1.5 SD excluded	$r(31) = 0.59$. $p < 0.001$, with a $R^2 = 0.35$	$r(30) = 0.59$, $p < 0.001$, with a $R^2 = 0.35$	$r(29) = 0.62$, $p < 0.001$, with a $R^2 = 0.38$
> 1 SD excluded	$r(25) = 0.73$, $p < 0.001$, with a $R^2 = 0.53$	$r(27) = 0.61$, $p < 0.001$, with a $R^2 = 0.37$	$r(26) = 0.76$, $p < 0.001$, with a $R^2 = 0.58$

As a consequence, further analysis was conducted of the datapoints that were more than one standard deviation from the trendline in order to investigate why currently there is not a strong correlation between OCGM and MoRiLE. All assessments were identified from the initial dataset for each OCG that had a least one risk, harm or threat outlier at one or more standard deviations from the trendline. These are summarized in table 5. It can be seen that only OCG2 had a risk assessment pair more than 2 standard deviations from the trendline. This one outlying risk assessment pair (see Outliers column for OCG2) was one of two assessment pairs used in the correlation analysis (see N column for OCG2). These two risk assessment pairs were drawn from the initial dataset comprising 11 OCGM and 3 MoRiLE assessments for OCG2. It can be seen that more risk assessment pairs are excluded as the number of standard deviations from the trendline is reduced. At 1.5 standard deviations, 3 risk assessment pairs are excluded (for OCGs 2, 9 and 15). At 1 standard deviation, 9 risk assessment pairs are excluded covering 7 different OCGs.

Table 5: Number of outliers of the risk, harm and threat correlations at 1, 1.5 and 2 standard deviations from the trendline.

	Initial dataset	**Paired dataset**

OCG	OCGM	MoR-iLE	N	Out-liers	Risk outliers (number of SDs)			Harm outliers (number of SDs)			Threat outliers (number of SDs)		
					>1	>1.5	>2	>1	>1.5	>2	>1	>1.5	>2
2	11	3	2	1	1	1	1	1	1	1	1	1	1
3	2	1	1	1	0	0	0	0	0	0	1	1	0
5	11	6	3	1	1	0	0	1	0	0	0	0	0
6	7	2	2	2	2	0	0	0	0	0	2	2	0
8	5	6	4	1	1	0	0	1	1	0	0	0	0
9	5	6	4	2	2	1	0	1	1	0	2	1	0
11	2	2	1	1	0	0	0	1	0	0	1	0	0
12	15	6	1	1	1	0	0	1	0	0	0	0	0
15	3	5	2	1	1	1	0	1	1	1	1	0	0
Total	61	36	23	11	9	3	1	7	4	2	8	5	1

The complete dataset of assessments for each of these OCGs was analyzed to help contextualize and explain the outlying assessment. To assist visual comparison of the OCGM and MoRiLE assessments of the same OCG, the scores were transformed to z scores using the formula:

z-score = (score – mean score) / standard deviation

This normalizes the mean and standard deviation of the OCGM and MoRiLE scores, to assist visual comparison of the scores.

Figure 9 shows the risk z scores for OCG2. Day 0 is the date of the first assessment of OCG2 in the initial dataset. The outlying paired scores are the penultimate OCGM assessment and the first MoRiLE assessment. This represents the outlier at more than two standard deviations in figure 6. The OCGM assessment precedes the paired MoRiLE assessment by 22 days. The three MoRiLE assessments are largely consistent with the final OCGM assessment. This could suggest that there was a rapid decline in the risk posed by the OCG during those 22 days which was reflected in both sets of assessments. This pattern is also reflected in the harm and threat assessments of OCG2. A similar pattern was found in the assessments of OCG5 and OCG8, in which all assessments were consistent with a rapid decline in the risk posed by the OCG.

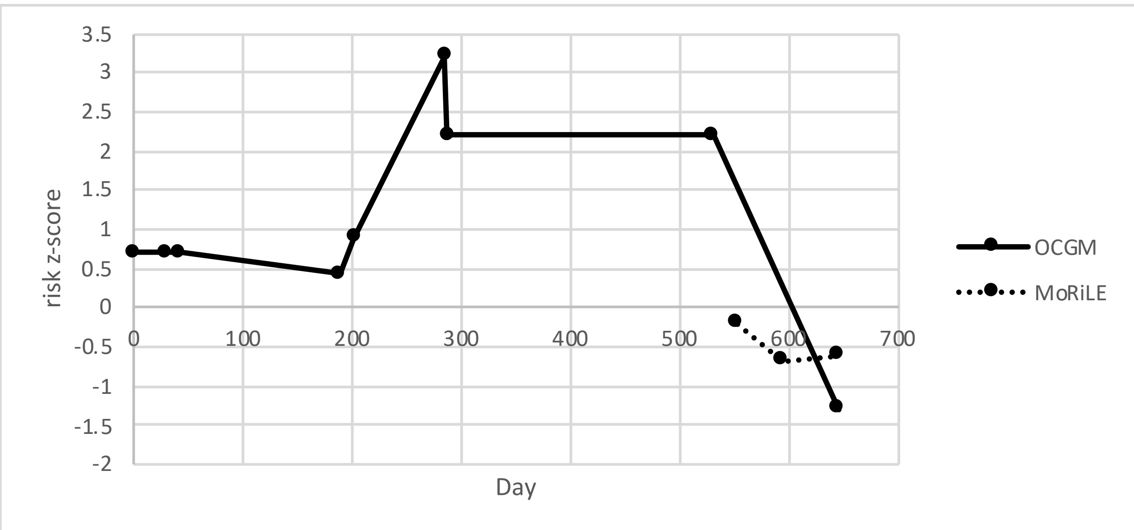

Figure 9: OCGM and MoRiLE risk z scores for OCG2. Days elapsed is shown on the x-axis.

The other outlying OCGs did not fit this pattern and could not be explained as there was no apparent convergence in the assessments. These OCGs with unexplained outliers all posed high levels of psychological and/or financial harm. The three outlier OCGs that could potentially be explained by a rapid change (i.e. OCGs 2, 5 and 8) all posed negligible or low levels of both psychological and financial harm. This may suggest a potentially lower level of agreement for OCGs that pose psychological and financial harm. It should be noticed that the assessments for these OCGs do not all deviate from the correlation in the same direction. Four of the unexplained outliers have higher OCGM scores, two have higher MoRiLE scores.

Analysis compared the outlier cases ($N = 11$) and non-outliers ($N = 25$) to test for differences across the OCGM and MoRiLE measures. Mann-Whitney tests were carried out to analyse differences between the outliers and non-outliers across the 17 OCGM and MoRiLE scores and sub-scores, using Bonferroni adjusted alpha levels (i.e. p value thresholds) of 0.003 (i.e. 0.05 divided by 17). OCGM harm was greater for outliers ($M = 47.45$, $SD = 28.91$) than for non-outliers ($M = 12.32$, $SD = 19.19$), $U = 40.0$, $p < 0.01$. No other significant differences were found between the outlier and non-outlier groups. This may suggest that reliability of the assessment methods can be lower for OCGs attracting higher OCGM harm scores. However, no other significant differences were found between the two sets of cases.

Predicting OCGM harm scores from MoRiLE harm factors

It has been shown that there is a moderate correlation between the OCGM and MoRiLE harm scores. Further analysis was conducted to investigate the extent to which the thematic factors that contribute to the MoRiLE harm score (i.e. physical, psychological, financial, community, environmental and geographic) can each predict the overall OCGM harm score. A multiple linear regression was calculated to predict the OCGM harm score based on the six MoRiLE harm factors (see table 6). A significant regression equation was found ($F(6, 29) = 3.687$, $p < 0.01$), with an R^2 of 0.433. Physical harm ($t = 2.609$, $p < 0.05$) and community harm ($t = 3.024$, $p < 0.01$) were significant predictors of the OCGM harm score. Psychological ($t = -1.979$, $p = 0.057$), financial ($t = 1.668$, $p = 0.106$), environmental ($t = 2.017$, $p = 0.053$) and geographic ($t = -0.865$, $p = 0.394$) harm factors were not significant predictors of the OCGM harm score. This suggests there is not a strong relationship between the underling MoRiLE harm themes and the process by which an analyst arrives at an OCGM harm score, with only physical harm and community harm having a predictive relationship to OCGM harm.

Table 6: Regression analysis summary for MoRiLE harm factors predicting the OCGM harm score.

MoRiLE harm factor	**B**	**SE B**	**ß**	**t**	**p**
Physical harm	15.340	5.880	0.450	2.609	0.014
Psychological harm	-6.560	3.315	-0.426	-1.979	0.057
Financial harm	1.196	0.717	0.265	1.668	0.106
Community harm	3.249	1.074	0.596	3.024	0.005
Environmental harm	24.695	12.241	0.314	2.017	0.053
Geographic scope	-0.623	0.720	-0.142	-0.865	0.394

Predicting OCGM threat scores from MoRiLE threat factors

It has already been noted that there was a moderate correlation between the OCGM and MoRiLE threat scores. A multiple linear regression was calculated to predict the OCGM threat score based on the five MoRiLE threat factors (see table 7). A significant regression equation was found ($F(5, 30) = 3.085$, $p < 0.05$), with an R^2 of 0.340. However, none of the individual elements namely, intent ($t = 1.420$, $p = 0.166$), capability ($t = 0.439$, $p = 0.664$), frequency ($t = 0.625$, $p < 0.537$), volume ($t = 0.253$, $p = 0.802$)

and victim vulnerability (t = -0.895, p = 0.378) were significant predictors of the OCGM threat score. This suggests there is no clear relationship between the underlying MoRiLE threat themes and the process by which an analyst would arrive at an OCGM threat score.

Table 7: Regression analysis summary for MoRiLE threat factors predicting the OCGM threat score.

MoRiLE threat factor	**B**	**SE B**	**ß**	**t**	**p**
Intent	11.891	8.372	0.363	1.420	0.166
Capability	2.729	6.218	0.085	0.439	0.664
Frequency	4.941	7.904	0.176	0.625	0.537
Volume	2.585	10.233	0.066	0.253	0.802
Victim vulnerability	-2.637	2.946	-.148	-0.895	0.378

Discussion

The risk assessment of OCGs can be carried out using either an attributed-focused or harm-focused method. Attribute-focused methods such as OCGM describe the OCG according to a set of attributes concerned, for example, with the criminality, intent and capability of the OCG. Harm-focused methods such as MoRiLE can be used to describe the OCG according to higher level thematic categories such as harm types. The analysis of risk assessment data suggests that OCGM and MoRiLE are in moderate agreement despite deriving risk assessment scores in different ways. However, a number of correlation outliers were identified. Some of these outliers could be interpreted as being due to a rapid shift in the risk posed by an OCG between successive risk assessments. Some outlying cases tended to be associated with OCGs causing psychological or financial harm or both. It could be speculated that certain modern criminalities such as cybercrime (Anderson et al, 2013) could fall within the category of criminalities that are associated more closely with psychological and financial harm. Such criminalities may pose particular challenges to the reliability of alternative methods of risk assessment.

Taken together, the outlying cases tended to be associated with OCGs causing a higher level of harm as assessed using OCGM. This could suggest that OCGs engaged in high levels of harm can potentially be more difficult to assess reliably across the two risk assessment methods. Regression analysis showed that the underlying risk and threat factors of MoRiLE can only to a limited extent predict the OCGM risk and threat scores. This further highlights the differences in process through the which the two types of risk assessment are made and the different OCG characteristics on which those processes draw.

As reliability might be lower with OCGs posing high levels of psychological or financial harm (as measured by MoRiLE) or high overall harm (as measured by OCGM), in such cases it might be beneficial to apply both OCGM and MoRiLE in combination and use both approaches to converge on an overall risk assessment for the OCG. Zoutendijk (2010) claims that many risk assessment methods are intersubjective: within the context of a particular method, consensus is reached as to what constitutes OCG risk and how it can be described and measured. However, it may not be possible to validate such a risk assessment method and there may be limited reliability (i.e. agreement) with alternative methods. Zoutendijk (2010) points out that this may require different countries to accept that they have different definitions of organised crime that cannot be easily compared. However, in the case of OCGM and MoRiLE, both methods are used in the same country and by the same policing organisations to map threats and prioritise resources. Therefore, just agreeing to differ would not be an appropriate position to take. However, if OCGM and MoRiLE were routinely used in combination to converge on a risk assessment, particularly in situations that tend to lead to low reliability, this process could help to increase consensus (i.e. intersubjectivity) among analysts, using the outcomes of both methods as sources in information.

Limitations

The analysis reported in this paper was carried out on OCGM and MoRiLE risk assessments of the same OCGs that could be sourced by the policing partner. It is not possible to determine the representativeness of the sourced assessments. For example, assessments carried out by certain analysts or for certain purposes may have been more easily accessible. Uncontrolled factors could also have affected the strength of correlation between the two sets of risk assessments. OCGM and MoRiLE scores for an OCG were paired if they were the closest possible match and no more than four months apart. However, it is unclear whether the paired assessments were carried out by the same analysts, for the same purpose and how and whether the intelligence on the OCG had changed between the two assessments. These factors could be expected to affect the strength of correlation between the two sets of assessments. As part of the research it was not possible to contact the analysts that had conducted the risk assessments. This study should therefore be considered as exploratory and indicating directions for further research. A follow-up study that could acquire feedback on the assessment process from the analysts involved as well as the assessment data, would facilitate a richer understanding of the assessment process and the tools that support it.

The analysis revealed a moderate correlation between the two assessment methods. However, even if there was a perfect positive correlation (i.e. an r value of 1) this would not necessarily imply that OCGM and MoRiLE assessments of the same OCG would lead to the same operational response. Investigating the relationship between MoRiLE and OCGM assessment outcomes and the subsequent operational responses is beyond the scope of this study. It is however of note that of the two methods, MoRiLE is the only one which incorporates an operational score, as such it is possible MoRiLE may be more useful, for example, to investigators tasked with considering resource requirements. Again, further research considering the needs and views of users would be worthwhile.

The focus of the current study was on investigating the reliability of OCGM and MoRiLE, i.e. the extent to which they are in agreement. The work did not consider the validity of the two methods, i.e. the extent to which the assessments reflect actual risk. The moderate correlation found between the two methods does not imply validity. Analysts applying similar knowledge and experience may be expected to reach similar risk assessments independent of the method used. However, their assessments may not necessarily reflect the genuine risk posed by the OCG. The validity of the risk assessment methods could be investigated by comparing the scores against objective measures such as future criminal convictions or the cost of harm caused. However, one confounding factor is that the outcome of risk assessment can be used set priorities. If an OCG was deemed to pose a high risk, police monitoring and interventions against OCG may affect the activities of the OCG and also what is known about them. These operational responses are therefore likely to change the reality that has been predicted. This is related to what Hamilton-Smith and Mackenzie (2010) describe as the 'problem of circularity' (p271). OCGs prioritised by the assessment method receive greater attention, revealing more about the criminality of the OCG, justifying and possibly further increasing the previous risk assessment. Therefore, determining the relative validity of OCGM and MoRiLE would be extremely challenging. However, as described previously, using both methods in combination would at least help to build intersubjectivity across the two methods, drawing on two different perspectives and sources of information.

Implications

This study, analysing paired OCGM and MoRiLE assessments, suggests that OCGM and MoRiLE are moderately in agreement with each other, despite arriving at risk assessments by rather different methods. Methodologically, this study also demonstrated how OCGM and MoRiLE scores can be matched into pairs for correlation, and how outliers of the correlation can be identified. These findings have potential implications for the future use of OCGM and MoRiLE for OCG risk assessment, which may also apply to other attribute-focused and harm-focused risk assessment methods.

First, as there is a correlation between OCGM and MoRiLE, the statistical approach of pairing and correlating scores across the two methods could be used routinely to compare data across the two methods and make refinements evidenced by the findings. For example, routine monitoring of the scores could be used to identify situations where there is relatively less agreement between the methods, for example, for particular types of criminality or harm. This finding could be used to inform revisions to the methods or to the supporting materials given to the analysts who undertake the risk assessment process. This finding could also be used to suggest when both methods should be used in combination in order that decision making can take into account the alternative assessments produced by the two methods.

Second, on the level of individual OCG risk assessments, statistical techniques could be used to identify outlying scores with respect to what is already known about the OCG. Given there is a level of agreement between the two methods, this could be done even if the new outlying assessment was done using a different method to all of the previous risk assessments. More broadly, this approach could be used to identify trends over time for OCG types, for example, whether there is a changing risk level for OCGs involved in certain types of criminality or associated with a particular geographical area. Practically however, as has been demonstrated in this analysis, OCGs are often involved in multiple criminalities and/or criminalities that cause multiple harms and that have a broad geographic scope. It can therefore be expected that any analysis of trends might need to consider multiple criminalities or harms across broad geographical areas.

Third, statistical analysis could also be extended to incorporate other sources of information. For example, given that the criminality and harm level of OCGs is being described using OCGM and MoRiLE, the resulting harm scores could be compared against external harm indices using measures such as sentencing guidelines (Sherman et al, 2016), court records (Bangs, 2016) or crime perception (Ignatans and Pease, 2016) as a source of additional information to inform OCG harm scoring.

Conclusion

This study compared two different methods for assessing the risk of OCGs, one attribute-focused (OCGM) and one harm-focused (MoRiLE). A moderate level of agreement was found between the two methods in terms of their assessment of risk, harm and threat. However, a number of correlation outliers were also identified. Some of these could potentially be accounted for by rapid changes in the risk posed by an OCG. Other outliers tended to be associated with OCGs posing a high level of psychological and/or financial harm. Regression analyses further highlighted the differences in how the two methods produce risk assessments and the attributes on which those assessments are based. What this research has highlighted is that in some instances, different decisions in relation to prioritisation could be made, depending on which risk assessment measure is used. As such is it suggested that the approach taken in this work could be used to inform revisions to the methods and their supporting materials, detect outlying risk assessments and potentially incorporate other sources of information such as harm indices. It is also suggested that using OCGM and MoRiLE in combination, particularly in situations where they tend to have lower agreement, could help to build broader intersubjectivity and consensus.

References

Albanese, J. S. (2008). 'Risk Assessment in Organized Crime: Developing a Market and Product-Based Model to Determine Threat Levels.' *Journal of Contemporary Criminal Justice*, **24** (3): 263-273.

Anderson, R., Barton, C., Böhme, R., Clayton, R., van Eeten, M. J. G., Levi, M., Moore, T. and Savage, S. (2013). 'Measuring the Cost of Cybercrime.' In: Böhme R. (ed.). *The Economics of Information Security and Privacy*. Springer, Berlin, Heidelberg.

Babyak, C., Matarazzo, A., Turner, J. and Wallace, M. (2009). 'Measuring Crime in Canada: Introducing the Crime Severity Index and Improvements to the Uniform Crime Reporting Survey.' https://www150.statcan.gc.ca/n1/en/catalogue/85-004-X

Bangs, M. (2016). 'Research outputs: developing a crime severity score for England & Wales using data on crimes recorded by the police.' https://www.gov.uk/government/statistics/research-outputs-developing-a-crime-severity-score-for-england-and-wales-using-data-on-crimes-recorded-by-the-police

Beken, V.T. (2004). 'Risky Business: A risk-based methodology to measure organised crime.' *Crime, Law & Social change*, **41**: 471-516.

Brown, J. (1998). 'Open-source information and the intelligence based decision model.' *Optimizing Open Source Information Conference*, 7-8 Oct 1998, Canberra.

Carr, R. Slothower, M. and Parkinson, J. (2017). 'Do Gang Injunctions Reduce Violent Crime? Four Tests in Merseyside, UK.' *Cambridge Journal of Evidence-Based Policing*, **1**: 195–210.

Crocker, R., Webb, S., Garner, S., Skidmore, M., Gill, M. and Graham, J. (2016). 'The impact of organised crime on local communities.' http://www.police-foundation.org.uk/2017/wp-content/uploads/2017/10/oc_in_local_communities_final.pdf

Dowden, C. (2017). MoRiLE programme Briefing Note (December 2017). https://polmorile.files.wordpress.com/2017/12/morile-programme-briefing-note-december-2017.pdf

Dubourg, R. and Prichard, S. (2007). 'Organised Crime: revenues, economic and social costs, and criminal assets available for seizure.' Home Office Online Report, 14(07).

Europol (2017). 'Serious and Organised Crime Threat Assessment.' https://www.europol.europa.eu/socta/2017

Hamilton-Smith, N. and Mackenzie, S. (2010). 'The geometry of shadows: a critical review of organised crime risk assessments.' *Policing & Society*, **20** (3): 257-279.

Ignatans, D. and Pease, K. (2016). 'Taking Crime Seriously: Playing the Weighting Game.' *Policing: A Journal of Policy and Practice*, **10** (3): 184–193.

Levi, M., (2016). 'The impacts of organised crime in the EU: some preliminary thoughts on measurement difficulties.' *Contemporary Social Science*, **11** (4): 392-402

Ratcliffe, J. H. (2015). 'Towards an Index for Harm-Focused Policing.' *Policing: A Journal of Policy and Practice*, **9** (2): 164– 183.

Sherman, L., Neyroud, P.W. and Neyroud, E. (2016). 'The Cambridge Harm Index: Measuring total harm from crime based on sentencing guidelines.' *Policing: A Journal of Policy and Practice*, **10** (3): 171–183.

Sullivan, C. and Su-Wuen, O. (2012). 'Justice Sector Seriousness Score: FAQs.' Wellington, NZ: Ministry of Justice.

Zoutendijk, J. (2010). 'Organised crime threat assessments: a critical review.' *Crime Law and Social Change*, **54** (1): 63–86.