Citation:
Shan-A-Khuda, M and Schreuders, C (2019) Understanding Cybercrime Victimisation : Modelling The Local Area Variations in Routinely Collected Cybercrime Police Data Using Latent Class Analysis. International Journal of Cyber Criminology, 13 (2). pp. 493-510. ISSN 0974-2891 DOI: https://doi.org/10.5281/zenodo.3708924

Link to Leeds Beckett Repository record:
http://eprints.leedsbeckett.ac.uk/6622/

Document Version:
Article

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please contact us and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

UNDERSTANDING CYBERCRIME VICTIMISATION: MODELLING THE LOCAL AREA VARIATIONS IN ROUTINELY COLLECTED CYBERCRIME POLICE DATA USING LATENT CLASS ANALYSIS

Mohammad Shan-A-Khuda & Z. Cliffe Schreuders[1]
Leeds Beckett University, United Kingdom

ABSTRACT

Numerous factors such as sociodemographic characteristics contribute to cybercrime victimisation. Previous research suggests that neighbourhood plays a role in cybercrime perpetration. However, despite the theoretical importance and particular interest to law enforcement agencies and policymakers, local area variations in cybercrime victimisation have rarely been examined. Drawing on data from recorded cybercrime incidents within one of the largest police forces in England from a three-year period with a victim dataset of 5,270 individuals enhanced by the Census data, this research untangles the relationships between demographics of cybercrime victims and their resident area characteristics. The research considers four types of cybercrime victimisation: 'Harassment/Unwanted Contact', 'Fraud/Theft/Handling', 'Sexual/Indecent Images' and 'other types of cybercrime' (classifications used by the participating police force). Latent Class Analysis (LCA) was applied to rigorously analyse the relationship among the four different types of cybercrime victimisation with victim demographics and resident area-level characteristics.

This research finds that each type of cybercrime yielded statistically distinct victim profiles. Vulnerabilities to cybercrime varied among male and female of different age groups, and importantly, the types of residential areas of the victims. Specifically, it is evident that females were much more likely to become cybercrime victims than males for two types of cybercrime: 'Harassment/Unwanted Contact', and 'Sexual/Indecent Images'. Vulnerabilities associated with these two types of cybercrime decreased with the increase of age. Cybercrime victims of 'Sexual/Indecent Images' were likely to be 5-14 year-olds living in areas with a higher number of Level 2, Level 4 qualifications and full-time students. Both males and females were vulnerable to 'Fraud/Theft/Handling' cybercrime and their resident areas had a higher number of full-time students, Level 4 qualifications and Asians. Finally, victims of 'other types of cybercrime' were most likely to be male and their resident areas had a high number of Asians and full-time students. Our work demonstrates that it is possible to apply statistical analysis to routinely collected police data to gain insight into the cybercrime victimisation that occurs across crime types in relation to demographics and area-level variations. These results provide valuable insights into policing cybercrime in England and beyond.

Keywords— Cybercrime, Victim Profiles, Area Variation, Policing Cybercrime, Evidence-based Policing, Latent Class Modelling

[1] Cybercrime and Security Innovation (CSI) Centre, Leeds Beckett University, Caedmon Hall, Headingley Campus, Leeds, LS6 3QS

1. INTRODUCTION

Cybercrime and related victimisation are increasing. Technology and digital devices have become ubiquitous in everyday life. Technology brings efficiency and effectiveness to a range of endeavours, including criminal behaviour, making new crime possible and enabling old crime to be conducted at unprecedented volume and speed. Cybercrime has recently surpassed, in volume, all other types of crime in the UK (NCA, & SCIG, 2016). The term cybercrime includes so-called cyber-dependent crime (or "pure cybercrime", which would not exist without said technology), and cyber-enabled and cyber-facilitated crime, which are the focus of this study, where technology has been used in the commission of a crime (UK Government, 2016).

Cybercrime can be described as being different from crime in the physical world in that the venue of cybercrime can be considered to have anti-spatial characteristics (that is, cybercrime can be committed from anywhere in the world, without physical access to the target of attack), often with a one-to-many relationship between criminal and victims (Brenner, 2004), and there is a high level of anonymity that is possible (Wall, 2007). Some of the common forms of cybercrime are sexual solicitation or harassment, identity theft, defamation, fraud, and phishing (Näsi et al., 2015).

While there is a body of research that has aimed to investigate the factors that relate to cybercrime victimisation, these studies often rely on self-reporting surveys focusing on college students, often with a small sample of actual cybercrime victimisation, and have typically not examined the complex relationship between types of victimisation and demographics of cybercrime victims and their resident area characteristics.

Our research is unique in that it benefits from access to police cybercrime records: 100% of our sample has been classified by police as cybercrime victimisation. This data includes further crime type categorisation, and we have combined this dataset with the Census data to accurately capture variations in neighbourhoods. Furthermore, our approach to analysis applies latent class analysis (LCA), a statistical analysis modelling technique which can be used to analyse sub-groups within a population, classifying cybercrime victimisation into mutually exclusive latent classes, which is an approach well suited to exploring the complex nature of the factors that relate to victimisation. We have applied LCA to identify combinations of factors that frequently appear together among victims of different types of cybercrime.

2. RELATED RESEARCH

As an increasing number of people fall victim to cybercrime, it is vital to study the factors that relate to victimisation. Both routine activity (Cohen and Felson, 1979) and lifestyle exposure theory (Hindelang et al., 1978), collectively referred to as lifestyle and routine activity theory (LRAT), focuses on the factors that make crime possible and contribute to victimisation risk. These theories have underpinned a body of research that seeks to understand the factors, such as victim demographics and behaviours that can lead to opportunities for criminal behaviour, and victimisation.

Quantitative studies have typically applied statistical analysis on self-reported victimisation surveys to explore contributing factors. In one such study, Holt and Bossler (2009) in

examining a specific form of cybercrime, online harassment, using a sample of college students, find that regular use of chat-rooms and other forms of computer-mediated communications were associated with victimisation risk; an explanation of which is that this can be attributed to an increased exposure to motivated offenders. In addition, committing computer deviance was found to increase the risk of online victimisation. The study also suggested that gender was an element in the process of applying LRAT to cybercrime victimisation. Overall, females were found to be more likely to be victimised due to being viewed as attractive targets, perhaps unrelated to computer-related behaviours and precautions.

Using structural equation modelling, Yucedal (2010) conducted a study applying LRAT to spyware and adware victimisation. Their study considered online activities and demographic variables (age, gender, marital status, education-elementary school to doctorate, race-white and non-white) in understanding effects on victimisation. Yucedal found that older people engaged in less online activities (holding other variables constant) which in turn exposed them to less cybercrime victimisation. In addition, males were found more likely to use the Internet for leisure activities such as visiting game websites and downloading materials from the Internet which in turn made them more vulnerable to malware. When education level increased individuals engaged in less leisure online activities. Finally, the study suggested that white people engaged in less leisure online activities than African-Americans, Hispanic and Asians when holding other variables constant.

Ngo and Paternoster (2011) applied the LRAT framework to investigate the role that individual and situational factors play in certain forms of cybercrime victimisation. Although individual and situational characteristics were not shown to consistently impact the likelihood of being victimized in cyberspace, the study found that both individual factors such as age, sex, race as well as situational factors such as exposure to motivated offenders, target suitability (such as communication with strangers), were associated with certain types of cybercrime victimisation. Using a sample of 295 USA college students, the study found that individual factors such as reported self-control had a statistically significant association with the victimisation of two of the seven studied cybercrime types: the probability of experiencing online harassment (i) by a stranger and (ii) non-stranger. The study also found that age, race, employment status and computer deviance such as looking at pornographic or obscene materials had a statistically significant association with cybercrime victimisation. This study also found that older individuals had a lower probability of experiencing malware, online harassment (by a stranger) or defamation.

Applying routine activity theory, Leukfeldt (2015) compared the risk factors for becoming a victim of two types of phishing: high-tech phishing (using malicious software) and low-tech phishing (using emails and telephone calls). Using data from a cybercrime victim survey with a sample of 10,316 in the Netherlands, and applying multivariate analyses, the study found that there were many similarities between two attacks as well as differences between high- and low-tech attacks. Two important aspects of the data preparation of this research were the integration of financial insights of the respondents using internal personal identification (RIN) and merging with external datasets. Consequently, the research used in total 30 variables split into four broad categories: (i) Sociodemographic such as gender, age, marital status, educational level (coded into eight categories from ‘no education’ to ‘university education’) and employment (12 hours per week or more); (ii) additional financial data (personal income, household income, value of financial assets, amount of savings); (iii) frequency of online activities, rated by respondents on a four-point scale; and finally (iv)

accessibility factors such as computer skills (a composite variable of knowledge about the operating system, Internet connection, web browser and antimalware in use) and risk awareness (a composite of ten variables such as "I open attachments of files from unknown senders" and "I use different passwords for different accounts").

The above studies all suggest that demographics, such as sex, ethnicity, age, and education levels, should be considered as factors that may affect cybercrime victimisation.

Previous research on cybercrime suggests that neighbourhood plays a role in cyber deviance. For example, using regression analysis Reinis (2016) found that (amongst many other contributing factors) neighbourhood integration (such as whether people in the neighbourhood can be trusted) and disorganization (such as whether there is a lot of graffiti in the neighbourhood) were associated with illegal downloading and hacking among adolescents attending high schools from 30 countries consisting of a sample size of 68,507 students. However, very few empirical studies have examined the relationship between cybercrime victimisation and the areas where the victims live.

In one of the few studies that have explored the relationship between victimisation and area variations, Näsi et al. (2015) used a combined four-country sample of 3,506 individuals (Finland, US, Germany and UK) to examine cybercrime victimisation among teenagers and young adults. The study reports only 6.5% of the total respondents had been a victim of cybercrime, and the differences in the cybercrime victimisation between countries were not statistically significant. However, male gender (perhaps due to the types of crime that were more common in this sample), younger age, immigrant background, urban residence, not living with parents, unemployment, and less active offline social life were found to be significant predictors of cybercrime victimisation.

The lack of empirical studies on cybercrime has been attributed to a lack of relevant data and the difficulty of collecting the data (Moitra, 2005). Bentaleb et al. (2015) describe some of the difficulties in estimating the population of victims of cybercrime, including the fact that some victims do not understand whether they were victims, or understand the definitions of cybercrime. Bentaleb goes on to suggest that cybercrime victimisation involves latent classes that cannot be measured directly.

Latent class analysis (LCA), is a statistical analysis modelling technique which provides opportunities for rigorous analysis that aims to understand sub-groupings in multivariate categorical data (McCutcheon, 1987). 'The primary aim of cluster-analytic statistical methods is to condense a set of classification objects into homogeneous groups (classes, cluster, types) – or to put it simply – to discover an empirical classification (taxonomy, typology)' (Bacher, 1996, translated from the original German). The classification is conducted in accordance with two criteria: (i) objects are to be grouped together should be as similar to one another as possible; and (ii) objects belong to different clusters should be as different from one another as possible. LCA can be beneficial in reducing the multiplicity of heterogeneous patterns of a complex social phenomenon that takes many different forms. LCA offers a statistical framework of inference based on the likelihood, probabilistic assignment of cases to classes and dealing correctly with the nature of indicator variables (Francis, 2016).

Bentaleb et al. (2015) present the application of LCA to test a model that identifies the proportion of youth population and proposes a measure of the degree of the risk of being cybercrime victim on social networks based on conditional probabilities. Using a sample of

"nearly 165 young internet users from six different regions in Morocco", this study tests models with different classes. The best model proposed has three latent classes. The classes from the LCA show that those with the highest risk are young people who ignore security measures, and those unable to assess the risk (54.5% and 46.2% respectively).

Particularly relevant to our research is the seminal work of Hirtenlehner et. al (2012), who applied LCA in reducing the multiplicity of heterogeneous patterns of stalking victimisation to a small number of distinct victimisation profiles. Their work developed a classification of stalking profiles that identified four distinct victimisation patterns using a data source from a survey of 311 Austrian stalking victims reported to the police. These patterns differ significantly in their determinants and their impact on the victim's well-being and quality of life. The determinants of victimisation profile take into account five measures: age of victims (in years), sex of victims (male and female), whether not the victim shared his or her home with a spouse or partner at the time the offences were occurring, sex of offender (male and female) and the type of pre-existing relationship between stalker and victim (stalking by an ex-partner and stalking by a non-ex-partner). Referring to the probabilistic assignment of cases to classes in LCA, Hirtenlehner et. al (2012) point out that 'class membership is determined by the maximum of the posterior probability of belonging to a particular class: A person is assigned to that class for which the membership probability is highest.'

Our research aims to overcome limitations in cybercrime victimisation studies, by applying LCA to a police dataset, to holistically explore the sociodemographic factors that relate to cybercrime victimisation.

3. METHODS

3.1. Data

Secure access to the cybercrime victim dataset was arranged under a Data Processing Contract (DPC) between Leeds Beckett University and the participating police force. We were granted access to a core dataset of police recorded cybercrime, with a sample size of 7364 victims, taking place from 2014 to 2016, within each of the districts. After deletion of cases with missing either sex or age, 4092 cybercrime victims were included in the LCA analysis using specialized software Latent Gold version 5.1.0.16259.

The percentages of cybercrime victims in the four categories were: Harassment/Unwanted contact (70.06%), Fraud/Theft/Handling (17.03%), Sexual/Indecent (12.29%), and other types of cybercrime (0.61%). The majority of victims were female (67.55%).

A unique characteristic of this victim dataset was that there was no other characteristic recorded about the victims such as race or ethnic classification, or occupation apart from age and sex. However, the Output Area Classifier code was made available and was used to link external data from the UK Census 2011 to the victim dataset.

This research thereby considered 36 different area variables grouped into four different categories: Ethnicity (White, Mixed Multiple Ethnic Group, Asian, Black, Other ethnicity), Qualification (No Qualification, level 1, level 2, level 3, level 4 and Other), National-Statistics Socio-Economic Classification (Higher Managerial and Administrative professional, Large Employers, Higher Professionals, Lower Managerial, Intermediate Occupations, Small Employers, Routine and Semi Routine, Never Worked, Long Term

Unemployment and Full-Time Students) and Occupation (Managers, Directors, Senior Officials, Professionals, Administrative, Secretarial, Skilled, Caring, Sales, Customer Service, Process Plant Operative and Elementary).

To facilitate the LCA, we have categorized each area-level measure into three categories: Low (below 25th percentile), Medium/Average (between 25th and 75th percentile) and High (above 75th percentile). For example, Table 1 and Table 2 shows the mean, median, mode and 25, 50 and 75th percentiles of the total number of Level 4 qualification (UK Government, 2019) and Economically Active Full-Time people living in the cybercrime victims' areas. Only four area-level variables (Asian, Level 2 qualification, Level 4 Qualification and Full-Time Student status) have been retained in the process of fitting the best model with the lowest value of statistical criteria after carefully examining the contribution of enunciating each of the 36 area-level variables in the latent class model.

Table 1: Level 4 Qualification Percentiles: Degree, Higher Degree, NVQ Level 4-5, HNC, HND, RSA Higher Diploma, BTEC Higher level, Foundation degree and Professional qualifications (UK Government, 2019).

Level 4 qualification		
N	Valid	4092
	Missing	0
Mean		50.31
Median		39.00
Mode		17.00
Percentiles	25	25.00
	50	39.00
	75	67.00

Table 2: Economically Active Employee Full Time Percentiles.

Economically (Full Time)	Active	Employee
N	Valid	4092
	Missing	0
Mean		83.24
Median		79.00
Mode		70.00
Percentiles	25	61.00
	50	79.00
	75	100.00

3.2. Strategy of analysis

The LCA was based on binary indicator outcome variables indicating whether a victim is subjected to each of the 4 different types of cybercrime: 'Harassment/Unwanted Contact', 'Fraud/Theft/Handling', 'Sexual/Indecent Images' and "Other types of cybercrime". We then profiled the latent segments in terms of demographics: gender, and age (in 8 groups); and area-level indicators: Ethnicity, Qualification, National-Statistics Socio-Economic Classification, and Occupation. Area-level indicators were considered inactive covariates.

Although latent variable(s) explain all of the associations among indicators, associations between covariates are not explained by latent variables (Kaplan, 2018).

Several Information Criterion such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and AIC3 (Adjusted Akaike Information Criterion) were examined to compare the models. The criterion for the final four-cluster model, representing the four cybercrime types, where all the demographic and area level measures have been considered, is presented in Table 3.

Table 3: Goodness of fit of the one- to four-class model.

Model fit statistics							
Model	LL	BIC(LL)	AIC(LL)	AIC3(LL)	Npar	L^2	df
One-class model	-6042.5845	12118.4362	12093.1690	12097.1690	4	5786.6479	206
Two-class model	-4145.0872	8423.2431	8322.1745	8338.1745	16	1991.6533	194
Three-class model	-3283.3698	6799.6098	6622.7397	6650.7397	28	268.2185	182
Four-class model	-3167.9461	**6668.5637**	**6415.8922**	**6455.8922**	40	**37.3710**	170

L^2 measures the lack of model fit (shared association among variables unexplained by the model) with the lower the value the better. The lower the value of AIC (LL), BIC (LL) and AIC3 (LL) also suggests that the four class model is the better fit of the model to the data. The cluster profiles for the best four class solution are presented in Table 4 in the next results section.

4. RESULTS

Table 4 shows the proportion of clusters and conditional probabilities of each of the variables to belong to a particular cluster with the demographic (gender, and grouped age) and four resident area level variables of the cybercrime victims under three broad umbrellas: Ethnicity (Asian), Qualifications (Level 2 and Level 4 qualifications) and Socio-Economic Classification (Full Time Students).

From Table 4, we can see that the largest cluster is the cluster 1 with 60% of victims falling into this cluster. Similarly, the second largest cluster is cluster 2 with 26% of victims. Cluster 3 has 13.4% and finally, the smallest cluster is cluster 4 with 0.005% of victims.

Table 4: Proportion of clusters (cluster size) and conditional probabilities of each of the variables to belong to a particular cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster Size	**0.6006**	**0.2598**	**0.1340**	**0.0056**
Gender of the victim				
Female	0.7343	0.7487	0.4363	0.3448
Male	0.2657	0.2513	0.5637	0.6552
Grouped Age (8 groups)				
5-14 years old	0.0610	0.5201	0.0085	0.1034
15-24 years old	0.3444	0.3178	0.2167	0.1379
25-34 years old	0.3043	0.0825	0.2351	0.3448
35-44 years old	0.1760	0.0431	0.2153	0.1724
45-54 years old	0.0900	0.0263	0.1544	0.2069
55-64 years old	0.0177	0.0095	0.0864	0.0000
65-74 years old	0.0057	0.0007	0.0510	0.0345
75-90 years old	0.0009	0.0000	0.0326	0.0000
Resident Area level variables (each variable is divided into three categories depending on the number of people living in that category)				
Asian				
Lower	0.2487	0.2389	0.1700	0.0690
Medium	0.5270	0.5157	0.4674	0.5862
Higher	0.2243	0.2454	0.3626	0.3448

Level 2 Qualification				
Lower	0.2385	0.2016	0.2875	0.3448
Medium	0.5289	0.5493	0.5227	0.4483
Higher	0.2325	0.2491	0.1898	0.2069
Level 4 Qualification				
Lower	0.2439	0.2272	0.1601	0.2414
Medium	0.5213	0.5215	0.5425	0.5172
Higher	0.2348	0.2513	0.2975	0.2414
Full Time Students				
Lower	0.2496	0.2096	0.1771	0.1034
Medium	0.5336	0.5500	0.5028	0.5172
Higher	0.2167	0.2403	0.3201	0.3793

Figure 1, Figure 2, Figure 3 and Figure 4 graphically represent victims' conditional probabilities of each of the demographic and resident area level variables to belong to cluster 1, 2, 3 and 4 respectively.

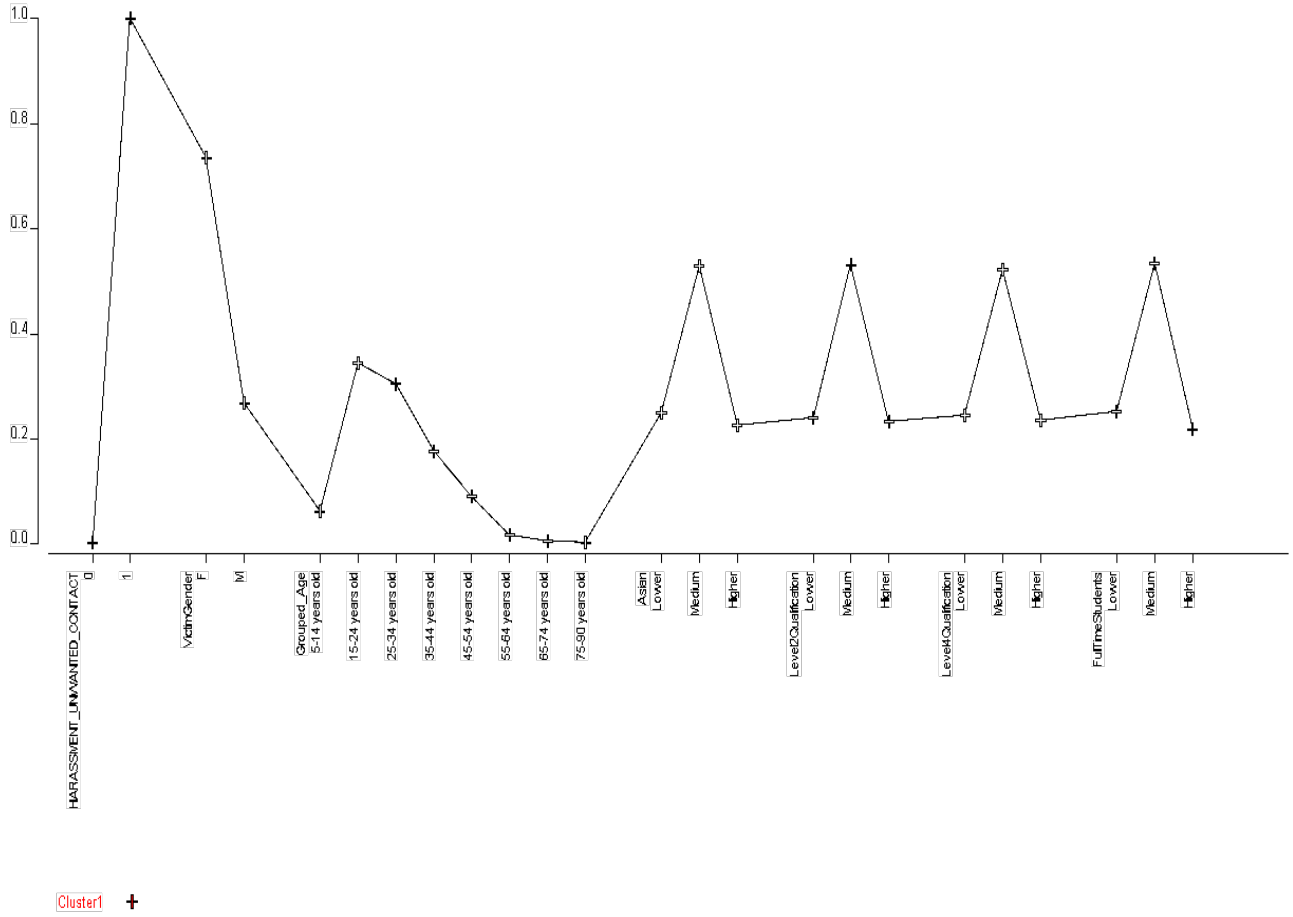

Figure 1: Conditional probabilities of demographic and area variables belonging to cluster 1.

From Figure 1, we can see that cluster 1 represents 'Harassment/Unwanted Contact' cybercrime victims. As illustrated in both Table 4 and in Figure 1, the conditional probabilities suggest that being in cluster 1, there is more than a 73% chance of being a female victim. In addition, there is 34%, 30% and 17% chances are that the age group of the victims will be 15-24, 25-34 and 35-44 years old respectively. As the age increases from 25, the chances of being a victim to this type of cybercrime decreases. There is higher than 50% probability that this particular type of cybercrime victims live in areas with a medium number of Asian, people with level 2, level 4 qualifications and full-time students. However, it is less likely that these type of cybercrime victims will live in areas with a higher number of level 4 qualifications and full-time students.

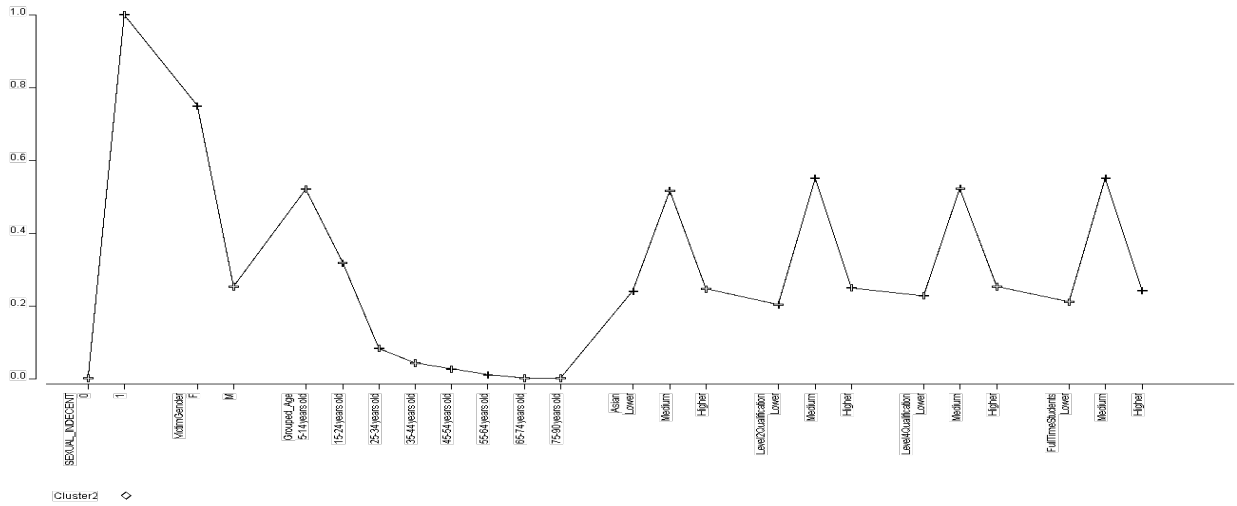

Figure 2: Conditional probabilities of demographic and area variables belonging to cluster 2.

From Figure 2, we can see that cluster 2 represents 'Sexual/Indecent Images' cybercrime victims. It is highly likely (with 75% probability) that for this type of cybercrime, victims will be female. There is more than 52% chance that for this type of cybercrime, victims are in the age group 5-14. The chances of being a victim of this type of cybercrime decreases with the increase of age. Compared to areas with a lower number of level 2, level 4 and full-time students, it is more likely that for this type of cybercrime victims live in areas with higher number of level 2, level 4 and full-time students.

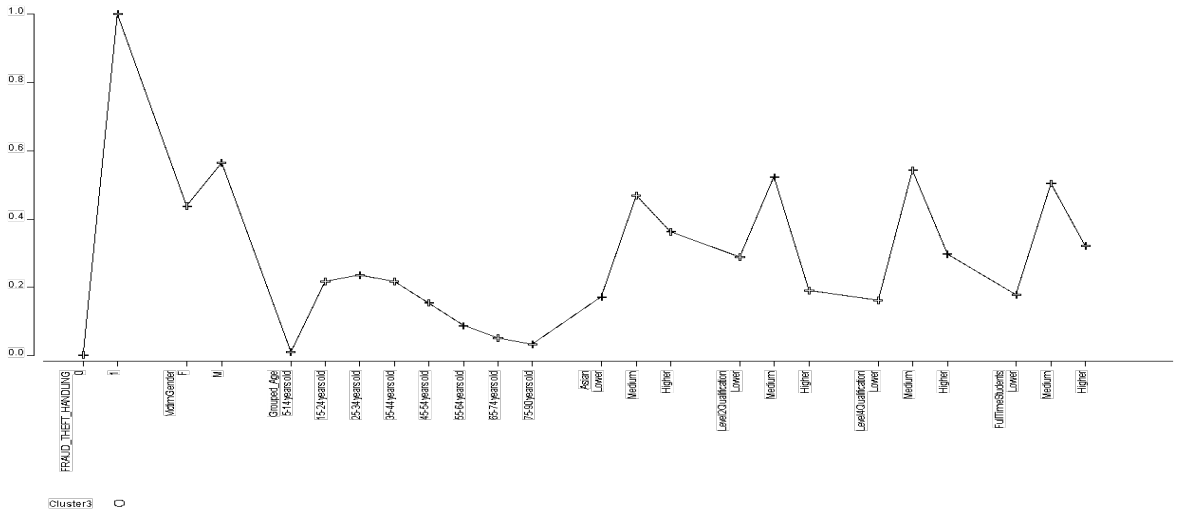

Figure 3: Conditional probabilities of demographic and area variables belonging to cluster 3.

From Figure 3, we can see that cluster 3 represents 'Fraud/Theft/Handling' cybercrime victims. For this type of cybercrime, victims are more likely to be male (56%) than female. As the age increases from 34, the chances of being victim to this type of cybercrime decreases. It is more likely that this type of cybercrime victims lives in areas with a higher number of Asian, level 4 qualification and full-time students.

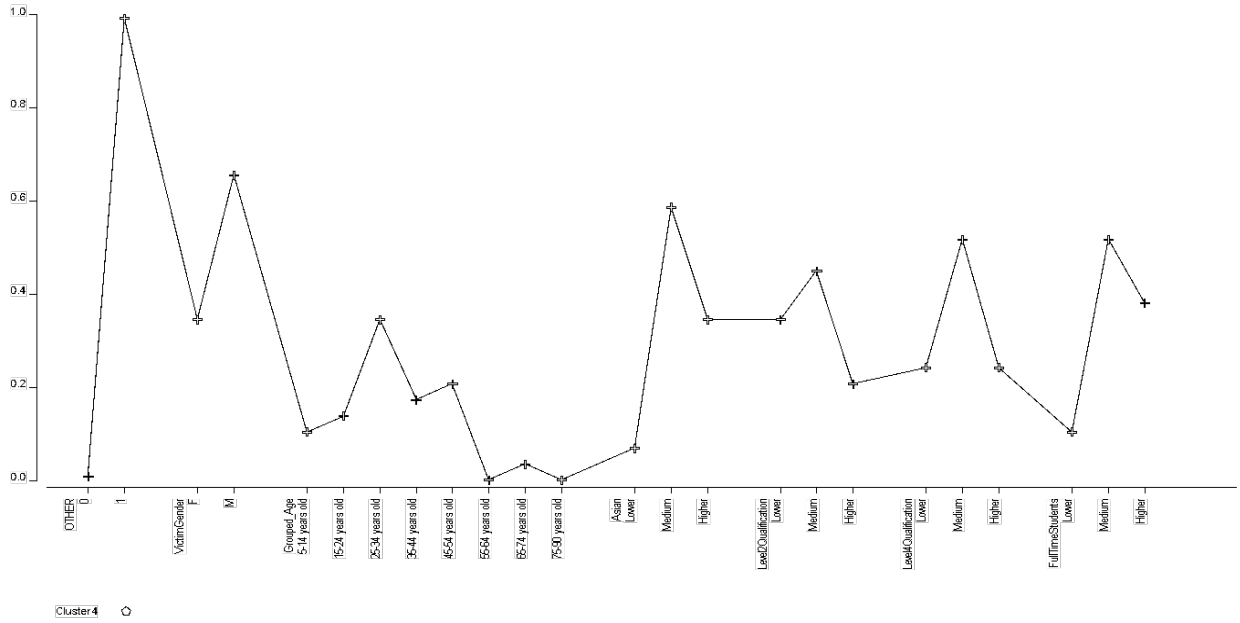

Figure 4: Conditional probabilities of demographic and area variables belonging to cluster 4.

From Figure 4, we can see that cluster 4 represents 'Other types of cybercrime' victims. It is much more likely that this type of cybercrime victims being male than female. The majority of the victims were males (66%). The probability of being victim to this type of cybercrime was the highest at 25-34 years. However, unlike the other three types of cybercrime victims, there is no decreasing pattern in probabilities as age increases. There was less chance of being in this type of cybercrime victims from the areas with a higher number of level 2 qualifications. There was a heightened chance of being in this 'Other types of cybercrime' victims from the areas with a higher number of Asian and full-time students.

5. DISCUSSION

A key contribution of our work is that it demonstrates it is possible to apply statistical analysis to routinely collected police data to gain a nuanced understanding of the cybercrime victimisation that occurs across each of the crime types in relation to demographics and area-level variances. This has the potential to assist police to target their response, and also deepen academic understanding of the phenomenon of cybercrime victimisation. The outputs from this research might be useful in evidence-based policing in the region and beyond.

Our analysis indicates that each of the types of cybercrime studied has statistically distinct victim profiles: resulting in a four-cluster set of profiles. As mentioned, within the dataset, cybercrime cases were categorised by police officers of the participating police force into four categories: Harassment/Unwanted Contact, Fraud/Theft/Handling, Sexual/Indecent, and Other types of cybercrime. Although this categorisation may simplify a more complex picture, the police force has found this categorisation useful for their internal monitoring and reporting use, and indeed only 0.61% were found by police to not fall into the three defined categories.

Prior studies on cybercrime victimisation have centred on both lifestyle exposure theory and routine activity theory (LRAT), and have indicated that several sociodemographic factors should be considered as factors that may be associated with victimisation.

Gender has been established as an important factor in LRAT that contributes to the likelihood of cybercrime victimisation. The LCA from this study confirms that in the British context, gender exerts a strong influence on cybercrime victimisation. Using a dataset of 4092 cybercrime victims within all of the districts of the participating police force, the current study finds that females were much more vulnerable to two types of cybercrime: 'Harassment/Unwanted Contact' and 'Sexual/Indecent Images'.

Previous studies on cybercrime victimisation have suggested that age as an individual factor has an impact on certain types of cybercrime victimisation. For example, Ngo and Paternoster (2011) found that older individuals had lesser odds of getting a computer virus, experiencing online harassment (by a stranger) or experiencing defamation. In line with previous studies, this study also finds that vulnerability towards two types of cybercrime: 'Harassment/Unwanted Contact' and 'Sexual/Indecent Images' decreased with the increase of age. However, there was no such decreasing vulnerability with older ages in the context of 'Fraud/Theft/Handling' and 'Other types of cybercrime'. Vulnerabilities associated with younger age groups from 5 to 24 have considerable impacts on the victims.

Vulnerabilities to 'Sexual/Indecent images' cybercrime at 5-14 years old (with a 52% probability of being female) carries a considerable degree of importance for law enforcement agencies, government and policymakers. According to Chief Constable Olivia Pinkney, the National Police Chief's Council Portfolio Lead for the policing of Children and Young People, a core role for policing is to protect the vulnerable in society. The National Policing Children and Young Persons Strategy 2013-2016 mentions that the age group of 18-24 year range is a key stage of development when the brain is still developing, independence is gaining, socialising activities are increasing (NPCC, 2015).

A key innovation of our study in studying the factors of cybercrime victimisation is the use of numerous Census based area-level measures to enhance the core police records victim dataset. The wide range of resident area-level measures considered reflects the types of households in those areas, which in turn associate to the type of neighbourhood of the resident areas of the cybercrime victims. Our study considered many potential sociodemographic variables to investigate whether they were associated with the types of cybercrime victimisation. This research considered 36 different area variables grouped into four different categories: Ethnicity, Qualification, National-Statistics Socio-Economic Classification, and Occupation. Four area-level variables were shown to contribute to the model: Asian, Level 2 qualification, Level 4 Qualification and Full-Time Student status. These four area-level variables have shown to contribute to the model based on the rigorous statistical criteria applied.

Our study confirms that the level of education and qualifications is a factor in cybercrime victimisation. Yucedal (2010) found that more educated people engaged in less online leisure activities, which may contribute to making them less vulnerable. In line with the previous study, our current study finds that 'Harassment/Unwanted Contact' cybercrime victims were more likely in areas with a lower number of level 2 and level 4 qualifications. As the number of level 2 and level 4 qualifications in an area goes up, the likelihood of becoming victim to this specific category of cybercrime was shown to decrease. In contrast, in the context of 'Sexual/indecent Images' cybercrime, the previous trend of the number of level 2 and level 4 qualifications reverses. The likelihood of becoming victim to this specific type of cybercrime increases in areas with higher levels of education. However, in the context of

'Fraud/Theft/Handling' cybercrime, it was evident that areas with a higher number of level 4 qualifications increased the vulnerability towards this particular type of cybercrime.

Previous studies on full-time students suggest that regular use of chat rooms and other forms of computer-mediated communications (Holt and Bossler, 2009), and lack of self-control (Ngo and Paternoster, 2011) were associated with risk of different cybercrime such as online harassment. In line with the previous studies, this study suggests that (with the exception of 'Harassment/Unwanted Contact') victims were more likely to be in areas with higher numbers of full time students (for 'Sexual/Indecent Images', 'Fraud/Theft/Handling' and 'Other types' of cybercrime).

Another important finding to emerge from this study is highlighting ethnicity as a factor in cybercrime victimisation. Previous studies, such as (Yucedal, 2010), found that ethnicity was associated with different online behaviours, and therefore exposure to victimisation. However, it is worth noting here that ethnicity in the study of Yucedal (2010) was a demographic variable of an individual cybercrime victim. Ethnicity in the current study is an area variable, which is the count of ethnic people (White, Mixed Multiple Ethnic Group, Asian, Arab and Other Ethnic Group) living in the areas of victims. This current study suggests that 'Fraud/Theft/Handling' and 'Other types of cybercrime' victims were more likely to live in areas with a higher number of Asian people. This does not account for the ethnicity of the victims themselves; although one possible explanation is that Asian communities are more vulnerable to these types of cybercrime. In line with the previous studies, this result suggests that ethnic differences have an impact on cybercrime victimisation. Thus, this finding has important implications for both criminological theory and those responsible for the development and implementation of social policy. In addition to the statistical results presented within this paper, police were also informed of specific hotspots of cybercrime victimisation that were identified during analysis, where specific areas within the region were found to have repeat victimisation, which were found to have high Asian populations.

6. LIMITATIONS AND FUTURE WORK

The main limitations of this research are related to the inherent properties of the routinely collected police data. Official cybercrime records do not contain a complete picture of victimisation. For instance, victims may not realise they have been victimised, and are not typically quick to report cybercrime incidents to law enforcement, often only after trying to resolve the situation themselves (HMIC, 2015). However, it could be argued that the official records provide a more complete picture than a self-selecting survey approach, as taken by many studies.

In interpreting our results, it is also important to note that the area-level variables in our study, such as ethnicity and level of education, cannot be attributed directly to the victims, but rather are attributes of the area the victims are located within. This disconnect may not impact the ability for police to target the areas that need specific kinds of attention; but may present an incomplete understanding on the underlying social causes. For example, it is not possible to ascertain quantitatively whether these factors increase the likelihood of perpetrators that happen to target neighbours (that may or may not match the sociodemographic characteristics), or whether the sociodemographic characteristics of the victims themselves are factors in victimisation. When the results of group-level/area-level

results are wrongly assumed to apply to the individual level, this is known as the 'ecological fallacy' (Robinson, 1950). Future research should also carefully consider the way results are interpreted.

Because of the lack of demographics about the victims in our core dataset, the current study could not include all the same sociodemographic variables employed in the study of Leukfeldt (2015) in comparing the risk factors for becoming a victim, such as marital status; frequency of online activities; or computer skills. Consequently, a likewise comparison of findings from this study to the previous study is not possible.

Our research is purely quantitative, working with existing datasets to gain insights into cybercrime victimisation to the extent possible using these datasets. Future qualitative research has the potential to contribute to a deeper understanding of the socioeconomic factors that play a role in cybercrime victimisation, and bring further insight into how quantitative results such as ours should be interpreted. Future research on cybercrime victimisation may benefit from a mixed methods approach.

A few limitations relate to the LCA approach. LCA assumes any association among the indicator or outcome variables is due to the latent variable being captured (Reid and Sullivan, 2009). This could result in erroneous conclusions if the association among the indicator variables were the results of chance or due to an unmeasured confounding variable (McCutcheon, 1987). Another limitation is that LCA also assumes that there is no within-class variation or that all members of a class have the same conditional item probabilities (Reid and Sullivan, 2009). Consequently, it is unclear whether individual differences are fully captured in the model (Lanza et al., 2007). Consequently, future research on cybercrime victims are encouraged to apply multilevel modelling (MLM) (Zhang and Reid, 2017). MLM is the accepted statistical technique for handling hierarchical data consisting of units grouped at different *levels* (Harvey, 2010). In MLM victims might be considered the level 1 units clustered or nested within the resident areas of the victims that may be the level 2 units. An analysis that models the way victims are grouped in areas has several advantages such as producing statistically correct estimates of regression coefficients, correct standard errors, confidence intervals and significance tests (Harvey, 2010). MLM would also make it possible to see whether some factors are better at accounting for or 'explaining' the variations in different cybercrime victimisation types, within different areas or districts. However, traditional single level models may suffice both for analysis and presentation for datasets depending on their structural complexity; multilevel models are tools to be used with care and understanding (Harvey, 2010).

7. CONCLUSION

The focus of this research has been in understanding the characteristics of cybercrime victims using their demographics and resident characteristics. LCA gives the probability of a case belonging to each group or cluster that are based on similarities; i.e. clusters of the cybercrime victims based on four different types of cybercrimes. Four distinct clusters are presented through LCA using different statistical criteria such as BIC (Bayesian Information Criterion) as a convenient method of choosing between models. In a brief, the crime type clusters are:

- 'Harassment/Unwanted Contact' cybercrime victims with 73% chances of being female. In addition, there is 34% and 30% chances are that the age group of the victims will be 15-24 and 25-34 years old respectively.

- ‘Sexual/Indecent Images’ cybercrime victims with 75% probability of being female and 52% belonging to age group between 5-14 years old.
- ‘Fraud/Theft/Handling’ cybercrime victims with more chances of being male (56%) than female (44%).
- ‘Other type of cybercrime’ victims with much more probability to belong to male than female.

This research has demonstrated that the official police record of cybercrime cases can be used to investigate area variation in the objective risks of cybercrime victimisation. This conclusion is also supported by some of the consistent findings between the current study and prior studies such as vulnerabilities of the female gender group to specific cybercrime types, and the impact of higher-level education towards heightened risks of cybercrime victimisation. Under the four categories of cybercrime presented in the dataset, this study presents the evidence base to associate cybercrime victimisation and the resident characteristics of the cybercrime victims. The identified area-level (societal) characteristics are important for policymakers and law enforcement agencies in reducing cybercrime victimisation. In this respect, future research on cybercrime may be benefited from LCA approach as employed in this research. The outputs from this research contribute to evidence-based policing through the development of the profiling of the victims that includes numerous sociodemographic factors.

This study is among the very few studies in understanding the area variations in cybercrime victimisation. Our findings have benefited one of the largest police forces in the UK, and have implications for other police forces within the UK as well as internationally in the context of ever-increasing cybercrime victimisation.

ACKNOWLEDGEMENTS

This work was supported by a Police Knowledge Fund grant, administered by the Home Office, College of Policing, and the Higher Education Funding Council for England (HEFCE).

The CARI Project is a large-scale collaboration between West Yorkshire Police and the cybercrime and Security Innovation Centre (CSI Centre) at Leeds Beckett University. The CARI Project aims to improve and incorporate an evidence-based approach into the policing of digital forensics and cybercrime investigations. An extensive needs assessment of UK policing and cybercrime and digital evidence was conducted to understand the current situation, and to identify needs across the force. The CARI Project also involved implementing a training and research programme that has impacted the capability of the digital forensics and cyber units within West Yorkshire Police to engage in research. This needs assessment and research training led to the development of a set of research proposals, which were scored and selected. Subsequently, academics and police staff co-produced 9 research and development workstreams: a framework for seizure, preservation and preservation of cloud evidence; automated forensic analysis; image linkage for victim identification and framework for image fingerprint management; automated grooming detection; frontline officer awareness development and decision support mobile app; assessment of methods of cyber training; an evaluation of the role of the Digital Media Investigator within WYP; and characteristics of victims of cybercrime. Each of these projects

were designed to address needs within law enforcement and outputs include evidence-based procedures, new capabilities such as software/algorithms, and actionable intelligence.

REFERENCES

Bacher, J. (1996). *Clusteranalyse: Anwendungsorientierte Einführung* (2. Bearb. u. erg.). München: Oldenbourg Wissenschaftsverlag.

Bentaleb, Y., Abarda, A., Mharzi, H., & Hajji, Said El. (2015). Probabilistic approach to estimate the risk of being a cybercrime victim, *9*, 6233–6240. http://dx.doi.org/10.12988/ams.2015.58559

Brenner, S. W. (2006). cybercrime jurisdiction. *Crime, Law and Social Change*, *46*(4), 189–206. https://doi.org/10.1007/s10611-007-9063-7

Brenner, S. W. (2004). Toward a Criminal Law for Cyberspace: Distributed Security.

Boston University Journal of Science & Technology Law, 10(2).

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588–608. https://doi.org/10.2307/2094589

Ford, J. K., MacCallum, R. C. and Tait, M. (1986) The Application of Exploratory Factor Analysis in Applied Psychology: A Critical Review and Analysis. Personnel Psychology, 39(2), pp.291-314.

Francis B (2016) Methods for Analysing Crime Data. Available at: https://www.ncrm.ac.uk/training/show.php?article=6205 (accessed 30 July 2018).

Harvey G (2010) *Multilevel Statistical Models, 4th Edition*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd. Available at: https://www.wiley.com/en-gb/Multilevel+Statistical+Models%2C+4th+Edition-p-9780470748657 (accessed 5 August 2018).

Hindelang, M. J., Gottfredson, M. R., & Garofalo, J. (1978). *Victims of personal crime: an empirical foundation for a theory of personal victimization*. Ballinger Pub. Co.

Hirtenlehner, H., Starzer, B., & Weber, C. (2012). A differential phenomenology of stalking: Using latent class analysis to identify different types of stalking victimization. *International Review of Victimology*, 0269758012446984. https://doi.org/10.1177/0269758012446984

HMIC, 2015. Real lives, real crimes: A study of digital crime and policing. HMIC (Her Majesty's Inspectorate of Constabulary), London.

Holt, T. J., & Bossler, A. M. (2009). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. *Deviant Behavior*, *30*(1), 1–25. https://doi.org/10.1080/01639620701876577

Kaplan D (2018) *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. Available at:

<https://uk.sagepub.com/en-gb/eur/the-sage-handbook-of-quantitative-methodology-for-the-social-sciences/book226672> (accessed 8 August 2018).

Lanza ST, Collins LM, Lemmon DR, et al. (2007) PROC LCA: A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 14(4): 671–694.

McCutcheon A (1987) Latent Class Analysis. London: SAGE.

Moitra, S. D. (2005). Developing Policies for cybercrime: Some Empirical Issues. European Journal of Crime, Criminal Law and Criminal Justice, 13(3), 435-464.

Näsi, M., Keipi, T., Räsänen, P., & Oksanen, A. (2015). cybercrime victimization among young people: a multi-nation study. *Journal of Scandinavian Studies in Criminology & Crime Prevention*, *16*(2), 203–210.

Ndubueze, P. N., Mazindu Igbo, E. U., & Okoye, U. O. (2013). Cyber Crime Victimization among Internet active Nigerians: An Analysis of Socio-Demographic Correlates. *International Journal of Criminal Justice Sciences*, *8*(2), 225.

NCA, & SCIG. 2016. Cyber Crime Assessment 2016, Need for a Stronger Law Enforcement and Business Partnership to Fight Cyber Crime. NCA Strategic Cyber Industry Group.

Ngo, F. T., & Paternoster, R. (2011). Cybercrime Victimization: An examination of Individual and Situational level factors. *International Journal of Cyber Criminology*, *5*(1), 773.

National Police Chief's Council (NPCC) (2015). National Strategy for the Policing of Children & Young People. Retrieved from https://www.npcc.police.uk/documents/edhr/2015/CYP%20Strategy%202015%202017%20August%202015.pdf

Reid, JA and Sullivan, CJ (2009). A Latent Class Typology of Juvenile Victims and Exploration of Risk Factors and Outcomes of Victimization. *Criminal Justice and Behavior* 36(10): 1001–1024. DOI: 10.1177/0093854809340621.

Reinis, U.. (2016). Cyber Deviance among Adolescents and the Role of Family, School, and Neighborhood: A Cross-National Study. *International Journal of Cyber Criminology*, *10*(2), 127.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. American Sociological Review, 15, pp. 351-357.

Tranmer, M. and Steel, D. G. (2001b) Ignoring a Level in a Multilevel Model: Evidence from UK Census Data. Environment and Planning A, 33(5), pp. 941-948.

UK Government, 2016. National Cyber Security Strategy 2016 to 2021. Retrieved from https://www.gov.uk/government/publications/national-cyber-security-strategy-2016-to-2021

UK Government, 2019. What qualification levels mean. Retrieved from https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels

Wall, D. S. (2007). *Cybercrime: The Transformation of Crime in the Information Age*. Cambridge, UK; Malden, MA USA: Polity Press.

Yucedal, B. (2010). *Victimization In Cyberspace: An Application Of Routine Activity And Lifestyle Exposure Theories*. Ph.D. thesis. Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=kent1279290984&disposition=inline

Zhang, J and Reid, Sa (2017). Multilevel Modelling. In: The International Encyclopedia of Communication Research Methods. American Cancer Society, pp. 1-13. DOI: 10.1002/9781118901731. IECRM 0160.