Accepted for publication in *Decision Support Systems*, version 2.1, 29 October 2015.

Causal inference for violence risk management and decision support in forensic psychiatry

Anthony Costa Constantinou[*,1,2], Mark Freestone[1,3], William Marsh[2], Jeremy Coid[1].

1. Violence Prevention Research Unit, Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine and Dentistry, Queen Mary, University of London, London, UK, EC1A 7BE.
2. Risk and Information Management Research Group, School of Electronic Engineering and Computer Science, Queen Mary, University of London, London, UK, E1 4NS.
3. Forensic Personality Disorder Service, East London NHS Foundation Trust, London, UK

THIS IS A PRE-PUBLICATION DRAFT OF THE FOLLOWING CITATION:

Constantinou, A. C., Freestone, M., Marsh, W., & Coid, J. (2015). Causal inference for violence risk management and decision support in Forensic Psychiatry. *Decision Support Systems*, 80: 42-55.

DOI: 10.1016/j.dss.2015.09.006
Corresponding author: Dr. Anthony Constantinou, E-mail: anthony@constantinou.info

[*] Corresponding author. E-mail address: anthony@constantinou.info

ABSTRACT

The purpose of Medium Secure Services (MSS) is to provide accommodation, support and treatment to individuals with enduring mental health problems who usually come into contact with the criminal justice system. These individuals are, therefore, believed to pose a risk of violence to themselves as well as to other individuals. Assessing and managing the risk of violence is considered to be a critical component for discharged decision making in MSS. Methods for violence risk assessment in this area of research are typically based on regression models or checklists with no statistical composition and which naturally demonstrate mediocre predictive performance and, more importantly, without providing genuine decision support. While Bayesian networks have become popular tools for decision support in the medical field over the last couple of decades, they have not been extensively studied in forensic psychiatry. In this paper we describe a decision support system using Bayesian networks, which is mainly parameterised based on questionnaire, interviewing and clinical assessment data, for violence risk assessment and risk management in patients discharged from MSS. The results demonstrate moderate to significant improvements in forecasting capability. More importantly, we demonstrate how decision support is improved over the well-established approaches in this area of research, primarily by incorporating causal interventions and taking advantage of the model's ability in answering complex probabilistic queries for unobserved variables.

Keywords: Bayesian networks, belief networks, causal interventions, criminology, forensic psychiatry, mental health, risk management.

1 INTRODUCTION

Adequate management of offenders released from *Medium Secure Services* (MSS) is crucial in preventing violent crime and ensuring efficient allocation of resources. To this end, many MSS in the UK and elsewhere make use of risk assessment tools in the ongoing management of patients. However, in this application domain, the current state-of-the-art is represented by regression-based models and checklists with no statistical composition. Forensic medical practitioners have remained unimpressed by the decision support offered by the current state-of-the-art in managing patients with serious mental illness problems, and have identified the need to examine new ways of modelling (Coid et al., 2015).

Bayesian networks (BNs), which are probabilistic graphical models based on causal relationships, are especially well-suited for decision making scenarios that require as to consider multiple pieces of uncertain evidence. Over the last couple of decades there has been a renewed interest in Bayesian inference, especially for real-world applications. This is because Bayesian inference, which used to be computationally intractable, now allow us to develop large-scale BN models using specialised software that takes advantage of efficient BN inference propagation algorithms (Pearl, 1988; Heckerman et al., 1995).

Since then, successful applications of BNs for decision support have been witnessed in various application domains. These include:

a) ***Law and forensics***: Fenton and Neil (2011) proposed the use of BNs as a tool for avoiding probabilistic fallacies in legal practice, which continue to occur despite that many of the fallacies have been well documented. Horman et al. (2014) used BNs as part of a novel approach to triage for digital forensics for collecting and reusing past digital forensic investigation information in order to highlight likely evidential areas on a suspect operating system.

b) ***Medical and biomedical informatics***: Yet et al. (2013a) presented a methodology for developing BN models that predict and reason with latent variables, in order to provide information that is useful for clinical decision makers, using a combination of expert knowledge and data, and in (Yet et al., 2013b) the authors described a decision support BN

system for assisting clinicians in making better decisions in Warfarin therapy management. Numerous other applications in biomedicine are covered in (Heckerman et al., 1992; Friedman et al., 2000; Lucas et al., 2000; 2004).

c) ***Safety*:** Naderpour et al. (2014) presented a decision support system to help with the management of abnormal situations in safety-critical environments and demonstrated, based on a case taken from US Chemical Safety reports, how the system provided support for operators in maintaining the risk of dynamic situations at acceptable levels. Qiu et al. (2014) proposed a BN for cascading crisis events, such as typhoons, rainstorms and floods, that provides the capability to analyse the chain reaction path of such an event and potential losses, with experimental results indicating that this BN-based method improved forecasting accuracy compared to existing classical methods.

d) ***Software development, Project Management, and Information Technology*:** Lauria and Duchessi (2006) developed a BN based on Information Technology implementations and demonstrated how the BN model can be incorporated into a decision support system to support *what-if* analysis. Hu et al. (2013) demonstrated how a BN model, that was learned from data but which considered expert causality constraints, was able to perform better, in terms of predicting project management risks, than many previously proposed well known algorithms and models. Yet et al. (2015) proposed a dynamic BN modelling framework for calculating the costs and benefits of a project over a specified time period, allowing for changing circumstances and trade-offs.

e) ***Sports prediction, betting and psychology***: Constantinou et al. (2012; 2013) demonstrated how an expert constructed BN model, that combined both data and expert knowledge, was able to outperform purely data-driven statistical models and generate profit against the gambling market. In sports psychology, Constantinou et al. (2014) employed a BN to infer referee bias diagnostically by examining whether relevant causal factors during a football match could explain referee decisions.

Decision support benefits from the use of Bayesian models have also been reported in other more specialised applications. For instance, Wang et al. (2011) proposed a hierarchical naïve Bayes model that improves existing identity matching techniques in terms of searching effectiveness, and Wu et al. (2015) developed a BN, as part of a framework for model integration and holistic modelling of socio-technical systems, and demonstrated decision support benefits based on an airport inbound passenger facilitation case study. Fenton and Neil (2012) illustrate how BNs can be applied to model knowledge in many more diverse fields.

However, there have been limited previous attempts in developing decision support systems using BNs in forensic psychiatry, as well as from questionnaire and interviewing data in general. Salini and Kenett (2009) acknowledged this by stating that BNs have been rarely used to analyse customer survey data. More specifically, these previous relevant attempts focused on analysing survey data for customer complaints and satisfaction (Blodgett & Anderson, 2000; Ronald et al., 2004; Salini & Kenett, 2009) and for marketing purposes (Ishino, 2014). Sebastiani and Ramoni (2001) also used survey data to extract general information from the British general household survey, which provides a continuous information on a range of social fields such as population, housing, education, employment, health and income. All of these previous studies have reported a number of advantages in using BNs for analysing this kind of data. Most notably, these include that BNs a) offer a rich and descriptive overview of the broader customer behaviour by providing insights into determinants and subsequent behavioural, b) provide a causal explanation using observable

variables within a single nonlinear multivariate model, c) provide the ability to conduct probabilistic inference for both prediction and diagnosis, and d) provide a graphical representation and outputs that be easily understood by professionals.

Despite the significant benefits demonstrated, BNs are still under-exploited in forensic psychiatry. Therefore, it was felt that causal BNs could improve on the current state-of-the-art. In (Constantinou et al., 2015b) we presented the first BN model for preventing violent re-offence in released prisoners with serious history of violence. This paper is an extension of that study, but which focuses on mentally ill patients and provides decision support for discharged decision making from MSS. The paper is organised as follows: Section 2 describes the data and methodology, Section 3 describes model validation and discusses the results, and Section 4 provides our concluding remarks and directions for future work.

2 DATA & METHODOLOGY

We make use of a dataset referred to as VoRAMSS (*The Validation of New Risk Assessment Instruments for Use with Patients Discharged from Medium Secure Services*; Doyle et al., 2014). The dataset consists of questionnaire, interviewing and assessment data from 386 patients, out of whom 343 are males and 43 are females. Interviews were performed at 6 and 12 months post-discharge. At 6 months post-discharge, the occurrences for general violence[†] and violent convictions are 13.73% and 2.33% respectively, while at 12 months (i.e. between 6 and 12 months after release) the respective occurrence rates are 11.40% and 3.12%. The cumulative rates (i.e. 0 to 12 months) for general violence and violent convictions are 22.28% and 5.18% respectively.

In addition to the VoRAMSS dataset mentioned above, we have also made use a small part of a second dataset which is referred to as the Prisoner Cohort Study (PCS) (Coid et al, 2009). This is because the PCS dataset provided information for a small number of model parameters that were considered important for decision analysis, but which the VoRAMSS dataset failed to capture (details in Section 2.2). However, the PCS dataset is somewhat different to the VoRAMSS dataset in the sense that it involves released prisoners, rather than patients discharged from MSS. However, many of those released prisoners also suffered from mental health problems (i.e. severe depression, anxiety, psychotic disorder). In an attempt to maintain relevant to the VoRAMSS dataset, we have restricted the cases considered by this second dataset to mentally ill individuals. The PCS dataset consists of questionnaire, interviewing and assessment data from 953 prisoners (before and after release), 778 males and 175 females, with a reconviction rate of 25.18% over a ~5 year period post-release. There were 594 cases of mentally ill individuals, and which were used to learn the causal relationships for the following model factors: a) anger management, b) drug misuse treatment, c) alcohol misuse treatment, d) cocaine dependence, e) cannabis dependence, f) stimulants dependence, and g) alcohol dependence.

All of this data that had been extracted from questionnaires, interviews and assessments of patients with a specialist was then combined with relevant patient data retrieved by the Police National Computer (PNC), which mainly consisted of criminal records. As a result, we were presented with a set of unstructured patient data that had been collected independently of the requirements of a BN model, with a large number of variables consisting of repetitive,

[†] The definition of General violence in this paper is identical to that of violence as defined in (Doyle et al., 2014) for the VoRAMSS dataset, which includes sexual assaults, assaultive acts that involved the use of a weapon; or threats that made with a weapon in hand as well as all the acts of battery, regardless of whether or not have resulted in injury.

redundant, and in many cases contradictory information. This meant that the initial format of the data was inappropriate for causal analysis. In order to make the data adequate for causal inference, we had to restructure the dataset.

Figure 1 presents a diagram which demonstrates the practices we had to use in order to move from the unstructured data into a BN model capable of simulating interventions for risk management decisions, indicating that expert knowledge played a crucial role at every step of the process. Expert knowledge was provided by two clinically active experts in forensic psychiatry (JC) and forensic psychology (MF), each with at least 8 years' experience in forensic mental health research, having published widely on: criminal justice outcomes (Fox & Freestone, 2008; Coid et al., 2011; Coid et al., 2013), psychopathy and personality disorder (Coid et al., 2012; Freestone et al., 2013), and mental illness (Coid et al., 2013). We discuss these development stages in turn in the subsections that follow.

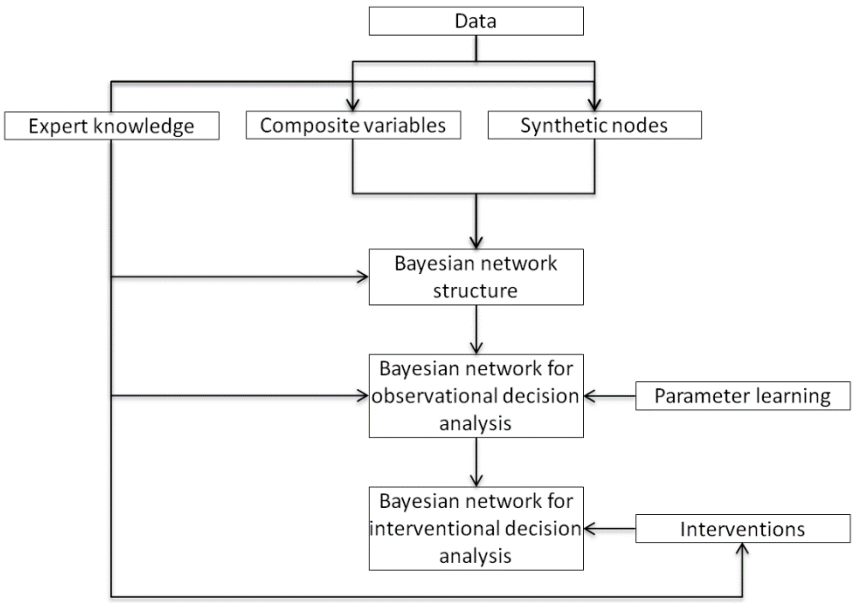

Figure 1. The process of developing the DSVM-MSS.

2.1. Constructing the Bayesian network structure

The causal structure of the BN model is solely based on expert knowledge. While the provisional BN structure was first drawn by hand, at the conceptual level, it was finalised only after all of the data management issues involving composite variables and synthetic nodes had been properly dealt with. We discuss each of them in turn below.

2.1.1. Composite variables

The first problem we had to deal with involves the formulation of composite information, based on a set of data variables and expert knowledge, that would deal with repetitive, redundant and contradictory information. For example, the model factor *General violence* represents a composite variable. It does not represent a single data variable, but rather a set of data variables and clinical judgments (i.e. violence reported by the clinician based on information mainly accounting to minor violent incidences) that, when combined, can provide a generalised indication with regards to *General violence*. Another example is *Personal*

resources[‡], which is a composite variable based on various sources of information including *Stable and suitable work, Effective coping skills* and other relevant data observations.

The challenge at this step primarily involves which sources of information to choose in order to inform a particular model factor, but also how to translate those sources into a unified single model factor. As an example, in informing *Personal Resources* we based the *learning* on five binary factors and introduced the following combinatorial rule:

> *if less than four of the selected factors indicate "No", then "Personal Resources="No", otherwise "Yes".*

Whereas for *Disinhibition*, which is based on four factors, we introduce an *OR* relationship between those four factor where *Disinhibition* would be *true* if any of those factors were *true*, otherwise *false*. This information is provided in Table B.4.

As shown above, the sources of information may include both data but also information that reflects the clinician's assessment. As a result, we found it impractical to derive a clear-cut method in determining how to inform the particular composite model variable and we, therefore, focused on expert judgments in determining the necessary data sources and ways of combining them into a unified model factor. The key idea from this part of the process is that, while it is far from perfect, it is certainly an improvement over throwing hundreds of variables into the network and expecting to form some sort of causal chain between them.

2.1.2. Synthetic nodes

We also made use of expert knowledge in introducing *synthetic* (or *definitional*) nodes within the causal structure. The synthetic variables are introduced for the purposes of a) reducing model dimensionality by combining different nodes together to reduce effects of combinatorial explosion (e.g. divorcing), and b) improving causal relationship between model variables.

Figure 2 presents, as an example, the elicitation of dependencies from experts for violence risk analysis in the first part of the diagram, whereas the second part demonstrates how the resulting complexity is managed by introducing three sensible synthetic variables (circled dashed nodes). Specifically, for this part of the model the experts suggested that the use of specific drugs (i.e. cannabis, cocaine, stimulants, and hazardous drinking) in conjunction with violent ideation and aggressive attitude (i.e. anger, hostility) that cannot be controlled (i.e. self-control) are believed to serve as causal risk factors for violence. However, if we were to model these 8 variables (variable states are presented in Table A.1) with direct links to violence (i.e. without introducing synthetic nodes) this would have resulted in a conditional probability table (CPT) for node *Violence* with $(3^2 \times 2^7) = 1152$ possible state combinations. Clearly, this would have been problematic given that the dataset considered for parameter learning only consists of just 386 data instances.

Reconstructing this part of the network, with the expertly defined synthetic nodes presented in Figure 2, not only reduced the combinatorial explosion by more than 97% (i.e. from 1152 down to $(3 \times 2^3) = 24$), and therefore allowed the formulation of more accurate CPTs, but also improved the causal relationship between factors for violence risk analysis. The expertly defined CPTs of all the synthetic variables introduced in this model are provided in Appendix B.

[‡] The node *Personal resources* is also modelled as a synthetic node into the BN. We cover synthetic nodes later in Section 2.1.3.

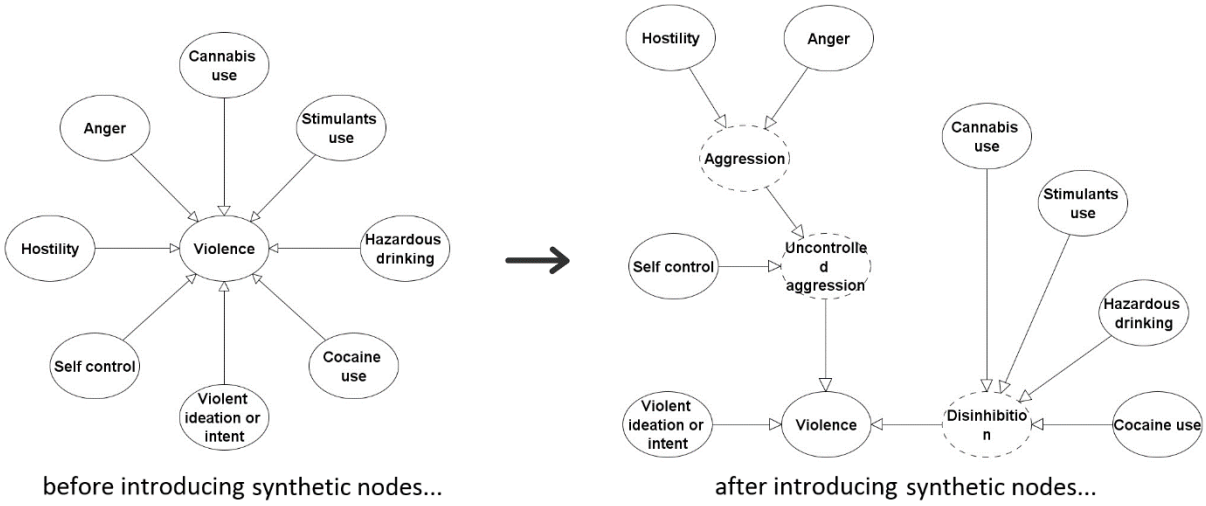

Figure 2. Synthetic nodes (circled dashed nodes) introduced for both reducing model dimensionality and improving the causal relationship between variables. This is a simplified model topology following the expert elicitation of dependencies for violence risk analysis.

2.2. Observational decision analysis

In this section we discuss the process of parameterising the model for observational decision analysis. The observational BN model is can be used to assess the risk of violence for a given mentally ill individual in the case of discharge. Most of the model parameters were learned from data, and just two model variables were learned with expert knowledge (excluding synthetic nodes).

These expert-driven variables are *Opiates dependence* and *Heroin dependence*, and which were considered to be important for violence risk analysis, but which data failed to capture. The expertly defined CPTs for these variables are provided in Appendix B. Various methods exist for expert probability elicitation, though most of them are not very different. For this task, we made use of probability scales and/or verbal anchors similar to those proposed in (van der Gaag et al., 1999; Renooij, 2001; van der Gaag et al., 2002). The experts were presented with the conditional probability tables of the other four substances (as learnt from the PCS dataset) in an attempt to assist them in providing rational conditional probabilistic judgments for heroin and opiates dependences.

All of the residual model factors (excluding synthetic nodes) were learned from data. In parameterising the CPTs of the data-driven factors, we had to rely on data which included a lot of missing value. As a result, we made use of the Expectation Maximisation (EM) algorithm, which is an iterative method for finding maximum likelihood estimates of parameters in models with unobserved latent variables (Lauritzen, 1995), and which represents the standard method for learning BN models from data with missing values.

2.3. Interventional decision analysis

In this section we demonstrate how we modified the resulting interventional BN from Section 2.2, into an interventional BN. The interventional BN is capable of performing for risk assessment, as in the observational BN, but also risk management by simulating interventions and examining their impact.

An intervention is an action which can be performed to manipulate the effect of some desirable future outcome which we would like to manage. In medical informatics, an intervention is represented by some treatment which can affect a patient's health outcome. In DSVM-MSS, *Anger management*, *Drug treatment* and *Alcohol treatment* represent uncertain (or imperfect) interventions. Much of the previous work, however, is focused on certain (or perfect) interventions (Pearl, 2000); implying that the intervention induces a specific state, rather than a distribution of states as in our case. Specifically, in DSVM-MSS an intervention answers questions such as: "*If a patient received his treatment/medication, what are the chances of him getting well?*".

An intervention is formulated as p(*E*|*I*), where *E* is the effect post-treatment and *I* is the intervention (Koller & Friedman, 2009). The intervention itself does not have parent nodes since we do not seek to explain the observation for treatment and hence, we must not reason backwards diagnostically. In order to satisfy this requirement under all circumstances, *graph surgery* (Pearl, 2000) must be performed on the observational BN, following parameter learning. By performing graph surgery, we modify the BN model such so that it becomes suitable for simulating interventional actions for the purposes of risk management and hence, the modified BN is described as the Interventional BN. Figure 3 demonstrates the process of introducing uncertain interventions in DSVM-MSS. Specifically, in the interventional case, a) any ancestor links entering the intervention (i.e. treatment) are removed, b) symptoms are manipulated by some intervention, and c) since we are dealing with uncertain interventions, the effectiveness of the interventions is determined by factors such as *Responsiveness to treatment* and *Motivation for treatment*. For more details on modelling interventions in Bayesian networks, including examples, see (Hagmayer et al., 2007). The intervention effectiveness rates have been taken by the model presented in Constantinou et al (2015b) and which are based on the PCS dataset.

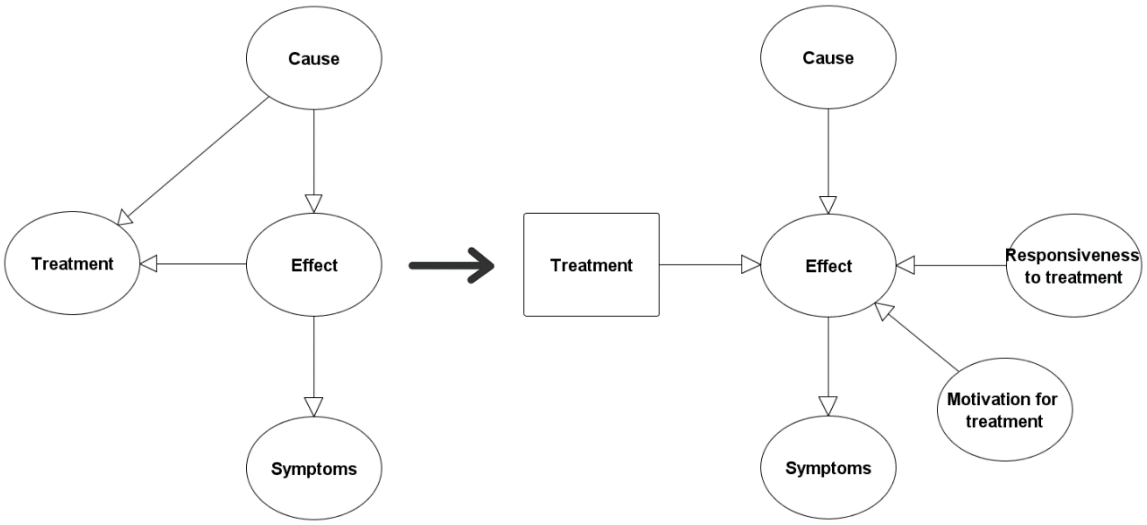

Figure 3. The process of introducing uncertain interventions in DSVM-MSS, and transforming the observational BN (left) into an interventional BN (right).

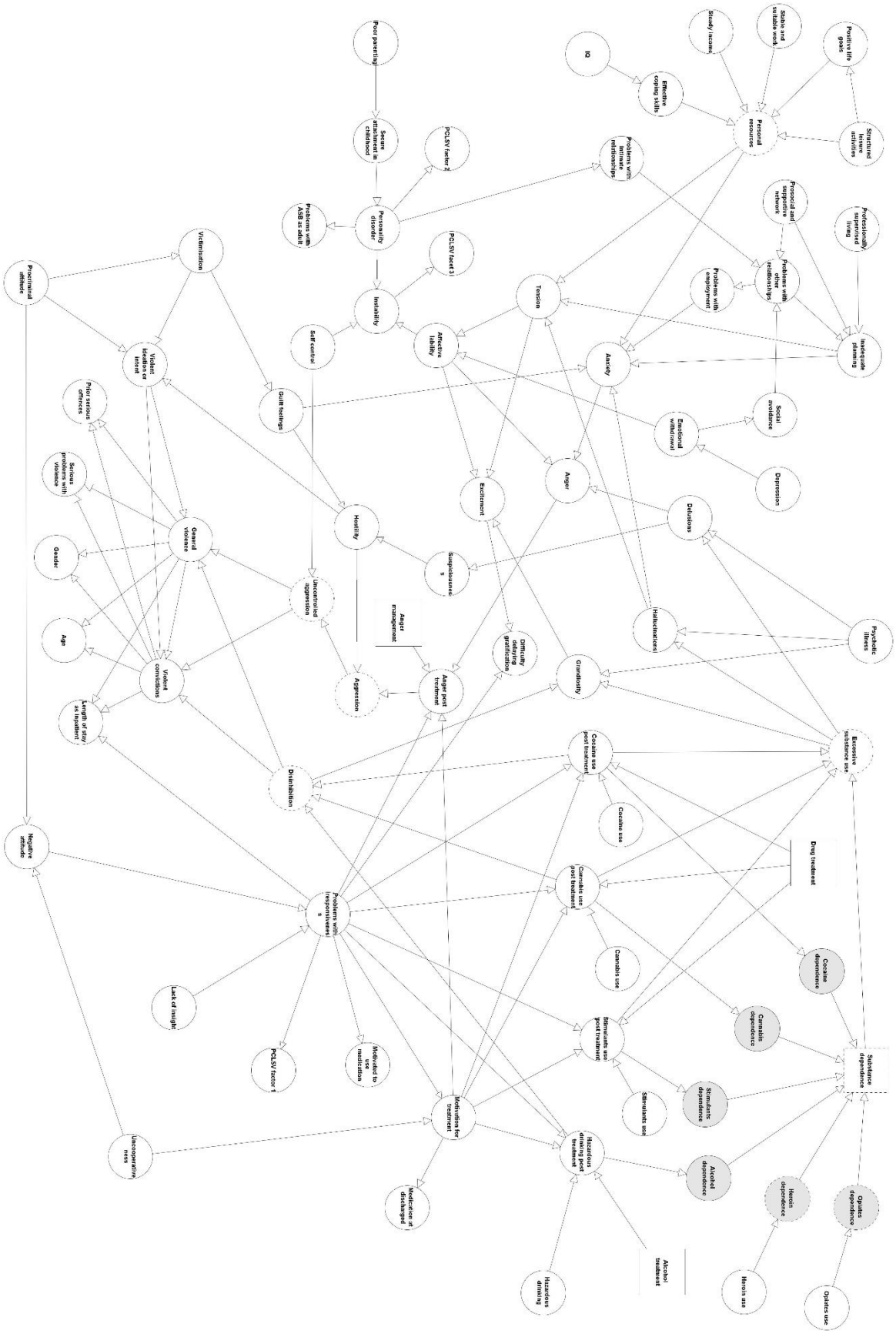

Figure 4. The complete Bayesian network model. *Circled solid nodes* are the variables learned with the main dataset, *circled dashed nodes* are the synthetic variables, *circled solid dark nodes* are the variables learned with the second dataset, and circled *dashed dark nodes* are the expertly defined variables, and square solid nodes are interventions.

The complete BN model is presented in Figure 4; the model variables which have been learned based on the main dataset (i.e. VoRAMSS) are represented by circled solid nodes, the variables which have been learned based on the second dataset (i.e. PCS) are represented by circled solid shaded nodes (excluding interventions), the variables whose CPTs are based on expert knowledge are represented by circled dashed shaded nodes, the synthetic variables are represented with circled dashed nodes, and interventions are represented by squared solid nodes. Figure 4 and Appendices A and B provide all the expert information required for someone to develop the model presented in this paper, with the VoRAMSS and PCS datasets.

3 MODEL VALIDATION & RESULTS

We provide two types of model validation, one which is data-driven and one which is expert-driven. Specifically, Section 3.1 assesses the forecasting capability of DSVM-MSS, in terms of predictive accuracy, by comparing the predictions generated by DSVM-MSS against those generated by models that are considered well-established in this application domain and thus, represent the current state-of-the-art. Further, Section 3.2 presents an expert-driven structural validation and assesses the capability of the model as a decision support tool, in comparison to the current state-of-the-art, for professionals who work in these areas.

3.1. Data-driven validation: Predictive accuracy

In assessing the predictive accuracy of the DSVM-MSS model we made use of the area under the curve (AUC) of a receiver operating characteristic (ROC). This is because the AUC of ROC is the preferred[§] measure of predictive or diagnostic accuracy in forensic psychology and psychiatry (Rice & Harris, 2005), and more than half of violence risk assessment validation studies report only the AUC (Singh, 2013). As a result, this allowed us to make direct comparisons of predictive accuracy against the current state-of-the-art.

The AUC is an evaluation metric for binary classification problems. The basic interpretation of this metric is that, given a random positive observation and a random negative observation, the AUC represents the proportion of the time the model correctly predicts the class. This independence of both base rate and selection ratio is appreciated in this application domain (Hanley & McNeil, 1982a, 1982b; Rice & Harris, 1995). The AUC score ranges from 0 to 1. A score of 0.5 indicates predictive capability no better than chance, whereas a score of 1 corresponds to a perfect predictive model (and vice versa).

First, we examine the predictive accuracy of the DSVM-MSS model. The AUCs are reported after performing *Leave-one-out* cross-validation (LOOCV), which involves using a single observation from the original sample as the *test* data, and the remaining observations as the *training* data over *n* iterations, such that every single data instance serves as out *test* data, where *n* is the total number of instances in the dataset. Table 1 presents the AUC scores achieved in predicting:

a) *General violence* (i.e. violence reported by the clinician, mainly amounting to minor violent incidences), and

[§] The AUC has also been subject to criticism on the basis that it provides an incomplete portrayal of predictive validity (Singh, 2013) and there is a debate in the literature on how the AUCs should be interpreted (Lobo et al., 2007). However, there is no other agreed measure for assessing violence risk assessment in this domain (Singh, 2013).

b) *Violent convictions* (obtained from a search of the Police National Computer).

The predictive assessment is provided for both at 6 and 12 months post-discharge (the dataset makes this possible because it contains interviewing and assessment data obtained at these time intervals). Specifically, Table 1 provides the following information:

a) Tests 1 and 2 provide the AUC scores at 6 months post-discharge, as predicted from relevant evidence up to the day of discharge, for general violence and violent convictions respectively.

b) Tests 3 and 4 provide the AUC scores at 12 months post-discharge, as predicted from relevant evidence up to the interviews performed at 6 months post-discharge, for general violence and violent convictions respectively.

c) Tests 5 and 6 simply represent the cumulative scores of the previous tests; i.e. the scores as generated throughout period 0 to 12 months post-discharge, for general violence and violent convictions respectively.

Table 1. AUC scores in predicting general violence and violent convictions.

Test	**Evidence period** (i.e. training data)	**Prediction period** (i.e. test data)	**Prediction**	**AUC**	**Lower** 95% CI	**Upper** 95% CI
1	At release	0-6 months post-discharge	General violence	0.691	0.619	0.764
2	At release	0-6 months post-discharge	Violent conviction	0.845	0.784	0.907
3	6 months after discharge	6-12 months post-discharge	General violence	0.730	0.655	0.805
4	6 months after discharge	6-12 months post-discharge	Violent conviction	0.774	0.591	0.957
5	Cumulative (test 1 & 3)	Cumulative (test 1 & 3)	General violence	0.708	0.656	0.761
6	Cumulative (test 2 & 4)	Cumulative (test 2 & 4)	Violent conviction	0.797	0.710	0.884

In both cases the model predicts violent convictions with higher accuracy than general violence. At 6 months, the model's capability in predicting violent convictions is considered to be significantly superior than predicting general violence (tests 1 and 2), given a *p*-value of 0.002, whereas this is not the case at 12 months (tests 3 and 4), given a *p*-value of 0.662. The *p*-value between the cumulative scores for general violence and violent convictions (tests 5 and 6) is 0.088.

Figure 5 examines the results for consistency by assessing whether significant discrepancies exist in predicting general violence and violent convictions between the two periods (i.e. 0 to 6 months against 6 to 12 months post-discharge). The ROC curves represent the true positive rate against the false positive, and provide measures for *sensitivity* (i.e. the proportion of positives which are correctly identified; in our case, correctly identifying violence) and *specificity* (the proportion of negative which are correctly identified; in our case, correctly identifying absence of violence). The diagonal line represents a random guess and hence, points above the line represent classification that is better than random, whereas points below the diagonal line represent classification that is worse than random. The light blue shaded area indicates 95% confidence interval AUC boundaries for the specified ROC curve. The results suggest that the model is rather consistent in predicting general violence with the AUC scores of 0.691 and 0.730 for the two specified periods, given that they demonstrate statistically insignificant difference in AUC assessment, with a *p*-value of 0.472. The same applies for violent convictions, with the AUC scores of 0.845 and 0.774 being statistically insignificant given a *p*-value of 0.469.

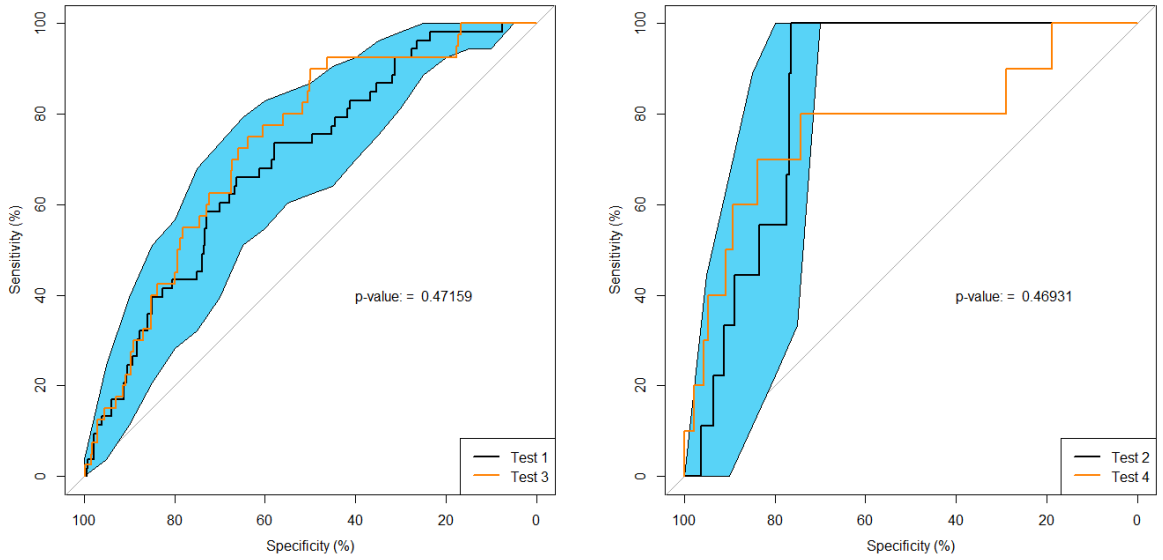

Figure 5. Resulting ROC curves based on tests 1 to 4 from Table 1. The light blue shaded area represents 95% confidence interval AUC boundaries.

Further, Appendix C presents supplementary information with regards to the model's capability in predicting self-control, hostility, anger, and violent ideation. These four factors were those elicited by the clinical experts as being causal for violent behaviour in our model[**]. When DSVM-MSS is employed with patients, this information will sometimes be unknown and therefore, it may be useful for clinicians to understand the capability of the model in predicting these four important factors. Table C.1 presents the AUC scores generated for each of these four factors, and Figure C.1 demonstrates the resulting ROC curves for each of the tests reported in Table C.1.

3.1.1. Predictive comparison against other models

In order to understand the predictive capability of the DSVM-MSS model, we need to compare the resulting AUC scores against those generated by models representing the current state-of-the-art in this domain. The following three well-established models have already been validated with the VoRAMSS in Doyle et al (2014):

1. ***HCR20 version 3*** (Douglas et al, 2013): a 20-item SPJ assessment of violence risk comprising ten static Historical (H) factors, such as previous violence; five dynamic Clinical (C) Factors relating to risk within forensic setting, such as impulsivity; and five dynamic Risk (R) factors relating to violence risk post-discharge, such as the existence of a personal support network. Items are scored on a three-point scale (0, 1, 2) depending on whether item is fully present and relevant to the patient (2); partially present and relevant (1); or absent (0).

2. ***SAPROF*** (de Vries Robbé et al, 2013): a 17-item checklist of static and dynamic protective factors; that is dynamic factors that are likely to 'protect' the patient from committing future violence. It includes items such as intelligence (static) and positive

[**] Anger, hostility and self-control serve as causal factors for violence, in our model, through the introduced synthetic nodes of *Aggression* and *Uncontrolled aggression* as presented earlier in Figure 2.

social network (dynamic), all of which are scored on the same trichotomous scale as the HCR20 above, and are grouped into *Internal* factors (relating to the patient's mental state, such as self-control); *Motivational* factors (relating to the patient's incentives to change, such as life goals); and *External* factors (relating to the patient's environment and social milieu, such as living circumstances).

3. ***Positive and Negative Symptom Scale*** (PANSS; Kay, Fiszbein & Opler, 1987) measures the severity of symptoms of mental illness. Symptoms can be positive (that is, outwardly displayed symptoms associated with psychosis, such as hallucinations or delusions); negative (relating to diminished volition and self-care in the patient); general (including non-specific symptoms such as depression); or relating to aggression in the patient.

The AUCs of the DSVM-MSS are compared with those for the total scale and sub-scales of each of the three predictors described above, as reported in Shaw et al (2013). We provide a detailed breakdown of the results in Tables 2 and 3. Specifically, Tables 2 and 3 report the results in predicting general violence and violent convictions respectively, at periods[††] 0 to 6 months and 0 to 12 months (i.e. cumulative) post-discharge. The results are as follows:

1. For general violence during 0 to 6 months post-discharge (Table 2), the DSVM-MSS is ranked 6th, out of 14 available predictions, in AUC score and performed significantly better against one model (out of 13; i.e. during test 11). No significant discrepancies between AUC scores had been observed for the residual tests. It is worth mentioning that the DSVM-MSS model demonstrated close to significant increase in performance against the model during test 2 (i.e. *p*-value 0.056).

2. For general violence during 0 to 12 months post-discharge (Table 2), the DSVM-MSS is ranked 1st in AUC score, and performed significantly better against three models (i.e. tests 2, 8 and 11). No significant discrepancies between AUC scores had been observed for the residual tests. It is worth mentioning that the DSVM-MSS model demonstrated close to significant increase in performance against the models during tests 4 and 12 (i.e. *p*-values of 0.053 and 0.065).

3. For violent convictions during 0 to 6 months post-discharge (Table 3), the DSVM-MSS is ranked 2nd in AUC score, and performed significantly better against five models (i.e. tests 6, 9, 10, 11 and 12). No significant discrepancies between AUC scores had been observed for the residual tests.

4. For violent convictions during 0 to 12 months post-discharge (Table 3), the DSVM-MSS is ranked 1st in AUC score, and performed significantly better against four models (i.e. tests 2, 6, 10 and 11). No significant discrepancies had been observed for the residual tests. It is worth mentioning that the DSVM-MSS model demonstrated close to significant increase in performance against the models during tests 3, 7, 9 and 12 (i.e. *p*-values of 0.063, 0.056, 0.056 and 0.065).

It is interesting to note that for all significant discrepancies in AUC score, the results were in favour of the DSVM-MSS model. While the AUC scores based on predictions for violent convictions look very promising, it should be noted that there is some uncertainty surrounding the results due to the low occurrence rate of violent convictions in the dataset. This concern is

[††] The previous published studies did not examine AUC scores in the period of 6 to 12 months post-discharge.

best illustrated by the respective confidence intervals, in AUC assessment, and resulting significance tests demonstrated above. Overall, however, the BN model appears to demonstrate moderate to significant improvements in violence risk assessment against the established clinical or regression-based models reported above, when employed with the same dataset.

Table 2. Significance tests between the DSVM-MSS and the regression models reported in (Shaw, 2013), which were also trained with the VoRAMSS dataset, for general violence at 6 and 12 months (cumulative) post-discharge.

Test	Model	AUC (0-6 months)	Lower 95% CI	Upper 95% CI	Significance	AUC (0-12 months)	Lower 95% CI	Upper 95% CI	Significance
1	HCRv3 Total	0.728	0.658	0.797	0.667	0.701	0.638	0.765	0.863
2	HCRv3 Historical	0.620	0.546	0.694	0.056	0.622	0.558	0.685	0.040
3	HCRv3 Clinical	0.746	0.679	0.813	0.390	0.705	0.644	0.767	0.942
4	HCRv3 Risk	0.663	0.589	0.738	0.331	0.626	0.561	0.691	0.053
5	SAPROF Total	0.764	0.705	0.823	0.169	0.692	0.631	0.753	0.700
6	SAPROF Internal	0.690	0.614	0.766	0.980	0.647	0.582	0.712	0.155
7	SAPROF Motivational	0.743	0.681	0.806	0.227	0.674	0.614	0.734	0.394
8	SAPROF External	0.658	0.587	0.729	0.516	0.621	0.555	0.686	0.040
9	PANSS Total	0.675	0.592	0.757	0.500	0.640	0.571	0.709	0.105
10	PANNS Positive	0.678	0.600	0.756	0.530	0.653	0.589	0.718	0.193
11	PANSS Negative	0.562	0.472	0.653	0.006	0.549	0.478	0.620	0.000
12	PANSS General	0.676	0.598	0.754	0.500	0.628	0.560	0.695	0.065
13	PANSS Aggression	0.716	0.634	0.798	0.877	0.680	0.613	0.747	0.514
-	BN model	0.691	0.619	0.764	N/A	0.708	0.656	0.761	N/A

Table 3. Significance tests between the DSVM-MSS and the regression models reported in (Shaw, 2013) which were also trained with the VoRAMSS dataset, for violent convictions at 6 and 12 months (cumulative) post-discharge.

Test	Model	AUC (0-6 months)	Lower 95% CI	Upper 95% CI	Significance	AUC (0-12 months)	Lower 95% CI	Upper 95% CI	Significance
1	HCRv3 Total	0.878	0.817	0.939	0.401	0.685	0.519	0.850	0.240
2	HCRv3 Historical	0.740	0.607	0.873	0.095	0.614	0.473	0.755	0.031
3	HCRv3 Clinical	0.768	0.686	0.850	0.174	0.659	0.543	0.775	0.063
4	HCRv3 Risk	0.835	0.735	0.934	0.834	0.656	0.478	0.834	0.164
5	SAPROF Total	0.814	0.704	0.923	0.548	0.674	0.517	0.832	0.178
6	SAPROF Internal	0.668	0.505	0.832	0.036	0.594	0.433	0.754	0.022
7	SAPROF Motivational	0.768	0.627	0.908	0.218	0.633	0.477	0.790	0.056
8	SAPROF External	0.826	0.724	0.929	0.728	0.685	0.528	0.842	0.193
9	PANSS Total	0.625	0.417	0.833	0.019	0.622	0.465	0.778	0.056
10	PANNS Positive	0.623	0.445	0.800	0.007	0.581	0.434	0.727	0.013
11	PANSS Negative	0.517	0.298	0.737	0.001	0.527	0.348	0.706	0.008
12	PANSS General	0.613	0.426	0.801	0.007	0.648	0.516	0.780	0.065
13	PANSS Aggression	0.716	0.518	0.915	0.139	0.659	0.493	0.824	0.149
-	BN model	0.845	0.784	0.907	N/A	0.797	0.710	0.884	N/A

3.2. Expert-driven validation: Causal structure & Decision support

This subsection covers two aspects of expert-driven model validation: a) examining that the causal structure of the model behaves rationally, and b) justifying the decision support provided by DSVM-MSS. We discuss these two expert-driven model validations in turn.

3.2.1. Structural validation with sensitivity analysis

Sensitivity analysis (SA) in BNs is a simple, but very useful, technique that analyses the impact of different model variables to a specified output variable. In our case, SA was used for rapid evaluation of the overall robustness of the BN model, as suggested in (Coupe & van der Gaag, 2000; van der Gaag & Renooij, 2001).

Specifically, this is done by assessing the impact selected model factors can have on a desired output variable. For example, Figure 6 demonstrates the sensitivity analysis for general violence between 0-6 months, and on the basis of the specified sensitivity variables. Since sensitivity analysis heavily depends on the causal structure of the BN, as well as on which model variables have been instantiated prior to performing the analysis, our experts were able to swiftly evaluate various such tornado graphs in an attempt to validate the structural integrity of the model. This was done by answering a series of simple questions, such as (examples from questions based on Figure 6):

a) "*From the nine influential factors considered, Age comes on top in terms of impact on General Violence, for the average mentally ill patient (e.g. Figure 5 assumes no instantiations and thus, represents the priors for the average individual). Is this reasonable?*".

b) "*When the average mentally ill patient is diagnosed with violent ideation or intend (VII), the patient's risk of becoming violent increases from 13.73% to 31.5% when VII=*"*true*", *and decreases from 13.73% to 9.8% when VII=*"*false*". *Is this reasonable?*".

If, for any reason, the experts identify some of the sensitivity values to be unreasonable, then this can be considered as an indication that there might be an error in one or more of the CPTs, or that part of the model's causal structure is inadequate.

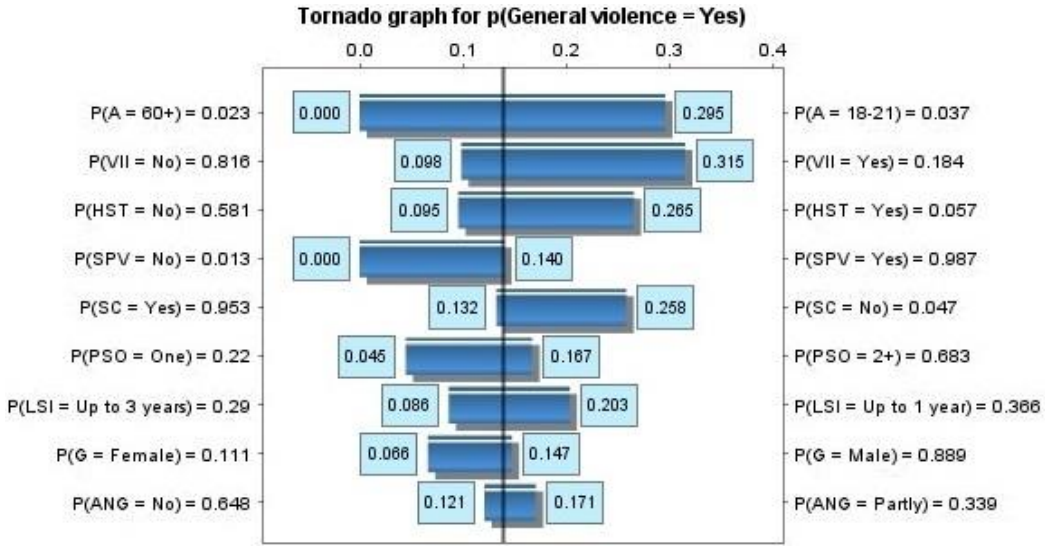

Figure 6. Sensitivity analysis for general violence, between 0 and 6 months after release, on the basis of the 9 specified sensitivity nodes and assuming no variable instantiations; where *A* is age, *ANG* is anger, *G* is gender, *HST* is hostility, *LSI* is length of stay as inpatient, *PSO* is prior serious offences, *SC* is self-control, *SPV* is serious problems with violence, and *VII* is violent ideation or intent. Probabilities next to the bars represent fluctuations for target node, whereas probabilities outside of the graph represent the prior probabilities for the specified state and variable.

Furthermore, in the same way SA can also be used to validate the inferences diagnostically. While diagnostic inference is harder for the experts to comprehend, compared to causal inference, it can still be useful for validating interventions. This is because in order to assess multiple factors against an intervention, the intervention has to be selected as the *target node*. However, the intervention always serves as a cause (i.e. treats symptom *S*) and hence, any inferences generated by the multiple sensitivity factors against the intervention only demonstrate diagnostic inference and ***not*** the actual effectiveness of the intervention. Figure 7, however, demonstrates why this can be useful for validation purposes.

Specifically, the target node here is *Drug treatment* against the eight predefined sensitivity factors. In this example, we have intentionally chosen some of the sensitivity factors to represent relevant symptoms (i.e. grandiosity, hallucinations, and delusions), relevant port-treatment effects (i.e. excessive substance use, disinhibition, cocaine post-treatment and cocaine dependence), and the risk for general violence which represents one of the outcomes for decision analysis. Figure 7 clearly demonstrates that all of the post-treatment effects share greater dependence against the intervention, in comparison to the relevant symptoms (which in also depend on other factors), whereas the risk for general violence hovers somewhere in the middle. Since this is the real-life behaviour one would expect, under the specified assumptions, we can conclude that *Drug treatment*, as well as the resulting structure between *Drug treatment* and the eight sensitivity factors considered, adequately simulate real-life expectations.

This process can be repeated for any factor of the model, for any set of sensitivity factors, and for any set of instantiations within the model. While the effort increases with large-scale BNs (as in our case; the BN consist of 80 factors), we found to be an extremely useful tool for this purpose.

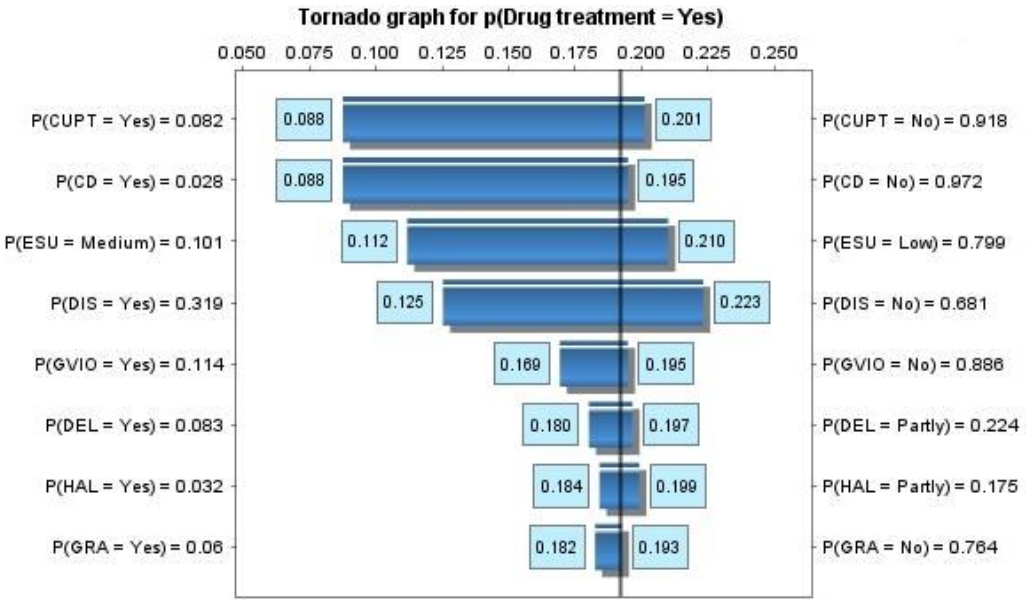

Figure 7. Sensitivity analysis[‡‡] for drug treatment, between 6 and 12 months after release, on the basis of the 8 specified sensitivity nodes and assuming no variable instantiations; where *CUPT* is cocaine use post-treatment, CD is cocaine dependence, ESU is excessive substance use, DIS is disinhibition, GVIO is general violence, DEL is delusions, HAL is hallucinations, and GRA is grandiosity. Probabilities next to the bars represent fluctuations for target node, whereas probabilities outside of the graph represent the prior probabilities for the specified state and variable.

[‡‡] Note that the results demonstrated here must assume that the intervention is uncertain (i.e. with prior probabilities as from learned data).

3.2.1 Decision support

Compared to the well-established predictors in this application domain, the DSVM-MSS model provides improved decision support that goes beyond predictive accuracy. Specifically, and as identified by the clinical experts and decision scientists, these decision support benefits are:

a) ***Risk management:*** One of the most important decision support features provided by the DSVM-MSS is its ability to simulate interventions for risk management. Unlike the other relevant predictors mentioned in this article, DSVM-MSS enables risk management professionals to prioritise interventions in an evidence-based fashion. Existing risk assessments such as the HCR20 or SAPROF may highlight individual risk factors, but provide no indication of the relative importance of individual factors to enable prioritisation of treatment or management. On the other hand, the BN model can not only illustrate the impact of each intervention on the desired output variable, but can also demonstrate visually how and at what degree an intervention influences the specified output for each individual, and this process can also be performed over combinations of interventions.

   In terms of decision support for the application domain, the BN model does this by examining whether the risk of an undesirable behaviour of a mentally ill patient can be managed to acceptable levels, as a result of one or more interventions, prior to determining discharge from MSS. This allows for analysis that answers complex clinical questions based on unobserved evidence; that is, "*how much could we expect to reduce the risk of violence for patient with profile A given intervention B?*". None of the current state-of-the-art models in this area of research are capable of simulating causal interventions.

b) ***Diagnostic inference:*** While current predictors are only capable of generating predictions from cause to effect, the BN model is also capable of inferring from observable effects to unobservable causes. In terms of decision support, this unique capability provides radically improved decision support since it enables extensive *what-if* analysis, in addition to *explaining away* unobserved variables. DSVM-MSS provides decision makers with the ability to investigate, for example, the reasons as to why a particular mentally ill patient behaved violently when the model was indicating otherwise, by inserting the relevant evidence into the model and allowing diagnostic inference to backpropagate to potential unobservable causes.

c) ***Management of missing information:*** The predictors which are currently established in this area require most, if not all, of their inputs to be entered in order to be accurate. In the case of DSVM-MSS, any such missing inputs are not ignored but rather inferred with revised beliefs based on the set of inputs which are known and thus, diminishing the limitation of not having a complete set of inputs.

   Moreover, in an extended version of this study we have demonstrated how the underlying principle of the *game-theoretic* technique *Value of Information* can be incorporated into these models to enhanced decision support by indicating whether a decision could be subject to amendments on the basis of some incomplete information within the model (i.e. when some inputs are missing in assessing a particular patient), and whether it would be worthwhile for the decision maker to seek further information prior to suggesting a decision (Constantinou et al., 2015a).

d) ***Expert knowledge and structural integrity:*** The current state-of-the-art predictors rely on classical methods and, in some cases, methods with no statistical composition. Hence, they typically only consider what data is available and assume linearity between risk factors since there is no causal or influential structure in place.

On the other hand, the BN framework is the most widely accepted modelling technique for incorporating expert knowledge along with relevant data. This has allowed us to incorporate expert knowledge for factors that are considered to be important for decision support but which historical data failed to capture, as well as to construct a structure with non-linear relationships between the variables of interest. The causal structure is intuitive and, by attempting to overcome the issues imposed by the unstructured data generated by the various questionnaires, interviews and clinical assessments, the model considers *what information we actually require*, rather than *what data we have available*. The model can retain its structure for future relevant studies, regardless how limited the new *training* dataset might be in terms of the number of variables.

4 CONCLUDING REMARKS & FUTURE WORK

The paper demonstrates a BN model, which we call DSVM-MSS, that can provide decision support to medical practitioners and professionals whose job involves determining whether a mentally ill patient is suitable for discharge from MSS. The motivation to develop DSVM-MSS arose from forensic psychiatrists and psychologists who have remained unimpressed by the decision support offered by the current state-of-the-art, which is represented by classical statistical methods and checklists with no statistical composition, and have, therefore, identified the need examine causal inference and the simulation of causal interventions. This paper is an extension of a recent study which focuses on the prevention of violent reoffending in released prisoners with serious background of violence (Constantinou et al., 2015b).

The results are based on both data-driven and expert-driven validations. In terms of the data-driven validation, DSVM-MSS demonstrates moderate to significant improvements in predictive accuracy when compared to the well-established models, that are employed with the same dataset, within this area of research. More importantly, however, and in terms of expert-driven validations, it is suggested that the DSVM-MSS is capable of improving decision support in a number of ways. These include the ability of the model to a) simulate causal interventions in an attempt to perform risk management for discharged decision making, b) perform diagnostic inference, c) manage missing information, and d) allow the incorporation of expert knowledge that allows the information to flow in a causally structured manner that is easily understood and appreciated by clinical experts. These benefits on predictive accuracy and decision support are discussed in greater detail in Section 3.

We have attempted to provide an adequate causal framework that captures the important properties of various aspects which could affect violent behaviour, such as mental illnesses, substance misuse, socioeconomic factors and personality disorder. As a result, the implications of this paper expand to both research domains; the forensic medical sciences and decision support systems. Specifically, when it comes to forensic medical sciences the paper attempts to direct medical practitioners and professionals into new ways of reasoning since the previous generation of models and predictors fail to deliver the decision support benefits that DSVM-MSS offers. In the case of decision support systems, the paper demonstrates how we managed to overcome the challenge of moving from a set of unstructured interviewing and clinical assessment data, which included repetitive, redundant and contradictory information, into a well-defined BN model that considers both data and expert knowledge for decision

support. This was achieved by focusing on what information we *really* require, rather than focusing on what data we have available, in order to meet the decision support objectives as identified by our domain experts. This adds to the limited previous attempts in developing decision support systems using BNs in forensic psychiatry, as well as from questionnaire, interviewing, assessment, and survey data in general.

While the process of BN model development requires an extensive iterative process between domain experts and decision scientists when modelling such highly complex real-world problems, BNs offer potential for transformative improvements. We believe that this type of modelling provides an important step forward for decision support within violence prevention research for individuals with enduring mental health problems. Further research and development should move beyond assessments of predictive accuracy, and into an evaluation of the efficacy of risk management decisions supported by the BN in 'real world' situations.

The problem addressed in this paper is typical of many critical real-world scenarios where decision makers require systems that go beyond regression and classification frameworks, especially in cases with limited or poorly structured data, and into improving decision support. The model presented in this paper will help in describing a method to systemise the development of BNs when the available information is based on questionnaire, interviewing or survey data, as well as to systemise the development of effective BNs for decision analysis in situations where there is limited data but access to expert knowledge. Both of these problems are being addressed in the BAYES-KNOWLEDGE project (Fenton, 2014).

ACKNOWLEDGEMENTS

The authors were supported by a Program Grant for Applied Research, program RP-PG-0407-10500, from The National Institute for Health Research UK (NIHR). We also acknowledge the anonymous reviewers whose suggestions have led to improvements in the paper.

APPENDIX A: *The variables considered by the DSVM-MSS model.*

Table A.1. Description of the model variables.

Variable No.	Node name	Node states	Dataset
1	IQ	*Low average/Average/High average*	VoRAMSS
2	Structured leisure activities	*No/Yes*	VoRAMSS
3	Stable and suitable work	*No/Yes*	VoRAMSS
4	Effective coping skills	*No/Yes*	VoRAMSS
5	Steady income	*No/Yes*	VoRAMSS
6	Positive life goals	*No/Yes*	VoRAMSS
7	Pro-social and supportive network	*No/Yes*	VoRAMSS
8	Professionally supervised living	*No/Yes*	VoRAMSS
9	Problems with intimate relationships	*No/Yes*	VoRAMSS
10	Problems with other relationships	*No/Yes*	VoRAMSS
11	Problems with employment	*No/Yes*	VoRAMSS
12	Social avoidance	*No/Partly/Yes*	VoRAMSS
13	Self-control	*No/Yes*	VoRAMSS
14	Inadequate planning	*No/Yes*	VoRAMSS

15	Personal resources	*Low/High*	Expert (Synthetic)
16	Delusions	*No/Partly/Yes*	VoRAMSS
17	Hallucinations	*No/Partly/Yes*	VoRAMSS
18	Anxiety	*No/Partly/Yes*	VoRAMSS
19	Depression	*No/Partly/Yes*	VoRAMSS
20	Grandiosity	*No/Partly/Yes*	VoRAMSS
21	Psychotic illness	*No/Yes*	VoRAMSS
22	Cannabis use	*No/Yes*	VoRAMSS
23	Cannabis use post treatment	*No/Yes*	VoRAMSS & PCS
24	Cocaine use	*No/Yes*	VoRAMSS
25	Cocaine use post treatment	*No/Yes*	VoRAMSS & PCS
26	Heroin use	*No/Yes*	VoRAMSS
27	Stimulants use	*No/Yes*	VoRAMSS
28	Stimulants use post treatment	*No/Yes*	VoRAMSS & PCS
29	Opiates use	*No/Yes*	VoRAMSS
30	Hazardous drinking	*No/Yes*	VoRAMSS
31	Alcohol treatment	*No/Yes*	PCS
32	Hazardous drinking post treatment	*No/Yes*	VoRAMSS & PCS
33	Drug treatment	*No/Yes*	PCS
34	Cannabis dependence	*No/Yes*	PCS
35	Cocaine dependence	*No/Yes*	PCS
36	Heroin dependence	*No/Yes*	Expert
37	Stimulants dependence	*No/Yes*	PCS
38	Opiates dependence	*No/Yes*	Expert
39	Alcohol dependence	*No/Yes*	PCS
40	Substance dependence	*No/Yes*	Expert (Synthetic)
41	Disinhibition	*No/Yes*	Expert (Synthetic)
42	Excessive substance use	*No/Medium/High*	Expert (Synthetic)
43	Personality disorder	*No/Yes*	VoRAMSS
44	PCLSV factor 1	*Low/Medium/High*	VoRAMSS
45	PCLSV factor 2	*Low/Medium/High*	VoRAMSS
46	PCLSV facet 3	*Low/Medium/High*	VoRAMSS
47	Poor parenting	*No/Yes*	VoRAMSS
48	Secure attachment in childhood	*No/Yes*	VoRAMSS
49	Instability	*No/Yes*	VoRAMSS
50	Problems with ASB as adult	*No/Yes*	VoRAMSS
51	Motivation for treatment	*No/Yes*	VoRAMSS
52	Motivated to use medication	*No/Yes*	VoRAMSS
53	Uncooperativeness	*No/Partly/Yes*	VoRAMSS
54	Negative attitude	*No/Yes*	VoRAMSS
55	Problems with responsiveness	*No/Yes*	VoRAMSS
56	Lack of insight	*No/Yes*	VoRAMSS
57	Medication at discharge	*No/Yes*	VoRAMSS
58	Tension	*No/Partly/Yes*	VoRAMSS
59	Guilt feelings	*No/Partly/Yes*	VoRAMSS
60	Affective lability	*No/Partly/Yes*	VoRAMSS
61	Anger	*No/Partly/Yes*	VoRAMSS
62	Anger management	*No/Yes*	PCS
63	Anger post treatment	*No/Partly/Yes*	VoRAMSS & PCS
64	Excitement	*No/Partly/Yes*	VoRAMSS
65	Suspiciousness	*No/Partly/Yes*	VoRAMSS
66	Hostility	*No/Partly/Yes*	VoRAMSS
67	Difficulty delaying gratification	*No/Partly/Yes*	VoRAMSS
68	Emotional withdrawal	*No/Partly/Yes*	VoRAMSS
69	Aggression	*Low/High/Very high*	Expert (Synthetic)

70	Uncontrolled aggression	*Low Aggression/High controlled/High uncontrolled*	Expert (Synthetic)
71	Gender	*Female/Male*	VoRAMSS
72	Age	*18-21/22-25/26-29/30-34/35-39/40-49/50-59/60+*	VoRAMSS
73	Length of stay as inpatient	*Up to 1 year/Up to 2 years/Up to 5 years/5+ years*	VoRAMSS
74	Pro-criminal attitude	*No/Yes*	VoRAMSS
75	Victimisation	*No/Yes*	VoRAMSS
76	Violent ideation or intend	*No/Yes*	VoRAMSS
77	Serious problems with violence	*No/Yes*	VoRAMSS
78	Prior serious offences	*None/One/2+*	VoRAMSS
79	General violence	*No/Yes*	VoRAMSS
80	Violent convictions	*No/Yes*	VoRAMSS

APPENDIX B: *Expertly defined CPTs*

Table B.1. Expertly defined CPT for synthetic node *Aggression*.

Anger	No			Partly			Yes		
Hostility	No	Partly	Yes	No	Partly	Yes	No	Partly	Yes
Low	1	1/2	1/2	1/2	0	0	1/2	0	0
High	0	1/2	0	1/2	1	1/2	0	1/2	0
Very high	0	0	1/2	0	0	1/2	1/2	1/2	1

Table B.2. Expertly defined CPT for synthetic node *Uncontrolled* a*ggression*.

Self-control	No			Yes		
Aggression	No	Partly	Yes	No	Partly	Yes
Low	1	0	0	1	0	0
High controlled	0	0	0	0	1	1
High uncontrolled	0	1	1	0	0	0

Table B.3. Expertly defined CPT for synthetic node *Excessive substance use*.

Cocaine use	No								Yes							
Cannabis use	No				Yes				No				Yes			
Stimulants use	No		Yes		No		Yes		No		Yes		No		Yes	
Substance dep.	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Low	1	0	1/2	0	1/2	0	1/3	0	1/2	0	1/3	0	1/3	0	0	0
Medium	0	0	1/2	0	1/2	0	2/3	0	1/2	0	2/3	0	2/3	0	1/3	0
High	0	1	0	1	0	1	0	1	0	1	0	1	0	1	2/3	1

Table B.4. Expertly defined CPTs for synthetic nodes *Disinhibition*, *Substance dependence*, and *Personal resources*.

Variable	Variable states	Parent nodes	Conditional definition
Disinhibition	No/Yes	Cocaine use, Stimulants use, Hazardous drinking, Cannabis use.	Disinhibition=No if all parent nodes=No, otherwise Yes (i.e. OR relationship between parent nodes)
Substance dep.	No/Yes	Cocaine dep., Stimulants dep., Cannabis dep., Alcohol dep., Heroin dep., Opiates dep.	Substance dep.=No if all parent nodes=No, otherwise Yes (i.e. OR relationship between parent nodes)
Personal resources	No/Yes	Steady income, Stable and suitable work, Positive life goals, Effective coping skills, Structured leisure activities	Personal resources=No if less than four parent nodes=No, otherwise Yes.

Table B.5. Expertly defined CPT for node *Opiates dependence*, given *Opiates use*.

Opiates use	No	Yes
No	1	0.84
Yes	0	0.16

Table B.6. Expertly defined CPT for node *Heroin dependence*, given *Heroin use*.

Heroin use	No	Yes
No	1	0.77
Yes	0	0.23

APPENDIX C: *Predictive assessment for causal factors for violence*

Table C.1 presents the AUC scores in predicting self-control, hostility, anger, and violent ideation. The results demonstrate that these factors are predicted with very high accuracy that is consistent over the two periods, in terms of AUC score. Self-control is the only factor which demonstrates some inconsistency between the AUC scores (i.e. without taking CIs into consideration) over the two periods. The high uncertainty generated at period 0-6 months in the AUC score for self-control might explain this inconsistency. Nevertheless, the *p*-values generated for each factor between periods do not demonstrate significant discrepancies between AUC scores and thus, the hypothesis for consistency for each factor between the two periods cannot be rejected; the *p*-values are: 0.172 for self-control, 0.469 for hostility, 0.090 for anger and 0.643 for violent ideation. In fact, the consistency hypothesis is closest to rejection for anger.

Table C.1. AUC scores in predicting self-control, hostility, anger, and violent ideation.

Test	**Evidence period**	**Prediction period**	**Predicted outcome**	**AUC** (95% CI)	**Lower** 95% CI	**Upper** 95% CI
1	At release	0-6 months after release	Self-control	0.638	0.419	0.857
2	At release	0-6 months after release	Hostility	0.787	0.669	0.905
3	At release	0-6 months after release	Anger	0.973	0.945	1.000
4	At release	0-6 months after release	Violent ideation	0.833	0.766	0.905
5	6 months after release	6-12 months after release	Self-control	0.810	0.697	0.922
6	6 months after release	6-12 months after release	Hostility	0.840	0.759	0.920
7	6 months after release	6-12 months after release	Anger	0.824	0.655	0.994
8	6 months after release	6-12 months after release	Violent ideation	0.811	0.748	0.875

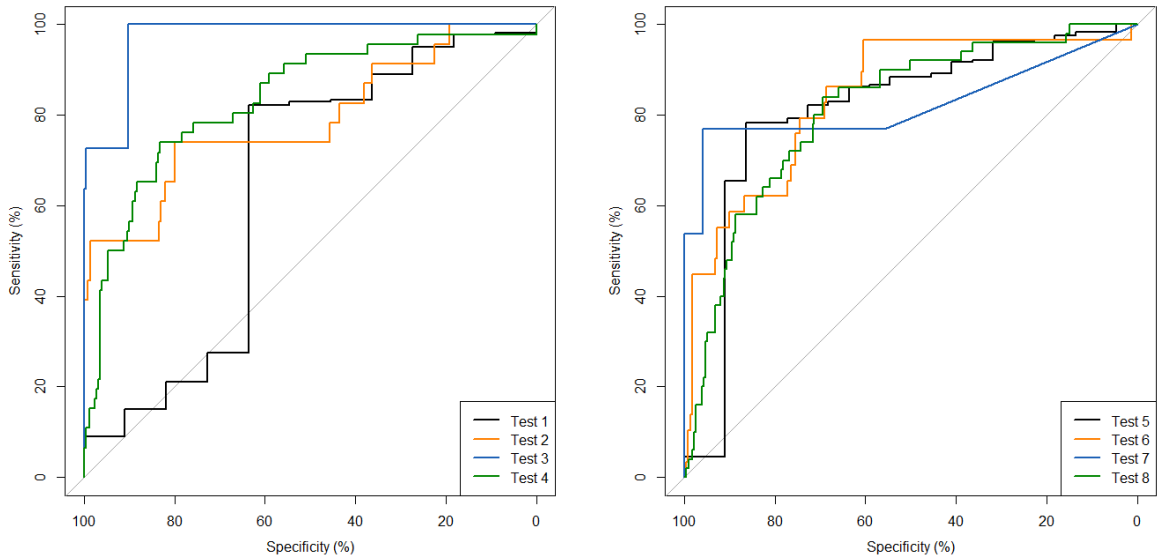

Figure C.1. Resulting ROC curves given the specified Tests 1 to 8, as reported in Table C.1.

REFERENCES

Blodgett, J. G., & Anderson, R. D. (2000). A Bayesian Network Model of the Consumer Complaint Process. *Journal of Service Research*, 2 (4): 321-338.

Coid, J. W., Ullrich, S., Kallis, C., Keers, R., Barker, D., Cowden, F., et al. (2013). The relationship between delusions and violence: findings from the East London first episode psychosis study. *JAMA Psychiatry, 70*(5), 465-471.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., et al. (2011). Most items in structured risk assessment instruments do not predict violence. *The Journal of Forensic Psychiatry & Psychology, 22*(1), 3-21.

Coid, J., Freestone, M. & Ullrich, S. (2012) Subtypes of psychopathy in the British household population: findings from the national household survey of psychiatric morbidity. *Social Psychiatry and Psychiatric Epidemiology*, 47(6): 879-891.

Coid, J. W., Ullrich, S., Kallis, C., Keers, R., Barker, D., Cowden, F., et al. (2013). The relationship between delusions and violence: findings from the East London first episode psychosis study. *JAMA Psychiatry, 70*(5), 465-471.

Coid, J.W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., et al. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Criminal Psychology*, 77, 337-348.10.1037/a0015155:

Coid, J. W., Ullrich, S., Kallis, C., Freestone, M., Gonzalez, R., Bui, L., Igoumenou, A., Constantinou, A. C., Fenton, N., Marsh, W., Yang, M., DeStavola, B., Hu, J., Shaw, J., Doyle, M., Archer-Power, L., Davoren, M., Osumili, B., McCrone, P., Barrett, K., Hindle, D., & Bebbington P. (2015). Improving Risk Management in Mental Health Services – A Multi-Methods Approach. The *National Institute for Health Research (NIHR)*, UK.

Constantinou, A. C., Fenton, N. E. & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36: 322, 339.

Constantinou, A. C., Fenton, N. E. & Neil, M. (2013). Profiting from an inefficient Association Football gambling market: Prediction, Risk and Uncertainty using Bayesian networks. *Knowledge-Based Systems*, 50: 60-86.

Constantinou, A. C., Fenton, N., & Pollock, L. J. H. (2014). Bayesian networks for unbiased assessment of referee bias in Association Football. *Psychology of Sport and Exercise*, 15(5): 538-547.

Constantinou, A. C., Yet, B., Fenton, N., Neil, M., & Marsh, W. (2015a). Value of Information analysis for Interventional and Counterfactual Bayesian networks in Forensic Medical Sciences. *Artificial Intelligence in Medicine*, Draft: http://constantinou.info/downloads/papers/VoIInterCounter.pdf

Constantinou, A. C., Freestone, M., Marshm W., Fenton, N., & Coid, J.W. (2015b) Risk assessment and risk management of violent re-offending among prisoners. *Expert Systems with Applications*, 42(21): 7511-7259. DOI: doi:10.1016/j.eswa.2015.05.025 .

Coupe, V. M. H., & van der Gaag, L. C. (2000). Sensitivity analysis: an aid for probability elicitation. *Knowledge Engineering Review*, 15: 215–32.

de Vries Robbé, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients. A retrospective validation study of the SAPROF. *International Journal of Forensic Mental Health*, 10, 178-186.10.1080/14999013.2011.600232.

Douglas KS, Hart SD, Webster CD, Belfrage, H. (2013). *HCR-20V3: Assessing Risk of Violence – User Guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Doyle, M., Coid, J., Archer-Power, L., Dewa, L., Hunter-Didrichsen, A. … & Shaw, J. (2014) Discharges to prison from medium secure psychiatric units in England and Wales. *British Journal of Psychiatry*, doi: 10.1192/bjp.bp.113.136622

Fenton, N.E. & Neil, M. (2011). Avoiding Legal Fallacies in Practice Using Bayesian Networks. *Australian Journal of Legal Philosophy*, 36: 114-151.

Fenton, N.E. & Neil, M. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press.

Fenton, N. (2014). Effective Bayesian Modelling with Knowledge Before Data. Retrieved June 14, 2015, from BAYES_KNOWLEDGE: https://www.eecs.qmul.ac.uk/~norman/projects/B_Knowledge.html

Fox, A. & Freestone, M (2008) An evaluation of the Good Lives and Discipline (GLAD) Programme at a UK DSPD Unit. *Prison Service Journal*, 151.

Freestone, M., Howard, R., Coid, J.W. & Ullrich, S. (2013) Adult antisocial syndrome with comorbid borderline pathology: association with antisocial and violent outcomes. *Personality and Mental Health*, 7(1): 11-21.

Friedman, N.F., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian Networks to Analyze Expression Data. *Journal of Computational Biology*, 7 (3/4): 601–620.

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In *Causal learning: Psychology, philosophy and computation*, ed. A. Gopnik & L. Schulz. Oxford University Press.

Hanley, J.A. & McNeil, B.J. (1982a). Maximum attainable discrimination and the utilization of radiologic examinations. *Journal of Chronic Disorders* 35: 601–611.

Hanley, J.A. & McNeil B.J. (1982b). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143: 29–36.

Heckerman, D., Breese, J., & Nathwani, B. (1992). Toward normative expert systems I: the PATHFINDER project. *Methods of Information Medicine*, 31: 90–105

Heckerman, D. E., Mamdani, A., & Wellman, M. P. (1995). Real-World Applications of Bayesian Networks - Introduction. *Communications of the ACM*, 38: 24-6.

Horsman, G., Laing, C., & Vickers, P. (2014). A case-based reasoning method for locating evidence during digital forensic device triage. *Decision Support Systems*, 61, 69-78.

Hu, Y., Zhang, X., Ngai, E. W. T., Cai, R., & Liu, M. (2013). Software project risk analysis using Bayesian networks with causality constraints. *Decision Support Systems*, 56: 439-449.

Ishino, Y. (2014). Knowledge Extraction of Consumers' Attitude and Behavior: A Case Study of Private Medical Insurance Policy in Japan. *The 8th International Conference on Knowledge Management in Organizations. Springer Proceedings in Complexity*, 425-438.

Kay, S.R., Fiszbein, A., & Opler, L.A. (1987). The Positive and Negative Syndrome Scale (Panss) for Schizophrenia. *Schizophrenia Bulletin*, 13, 261-276.

Koller, .D, & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Lauria, E. J. M., & Duchessi, P. J. (2006). A Bayesian Belief Network for IT implementation decision support. *Decision Support Systems*, 42: 1573-1588.

Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19: 191-201.

Lobo, J. M., Jimenez-Valverde, A., & Real, R. (2007). AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 1-7.

Lucas, P. J., van der Gaag, L. C., & Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial Intelligence in medicine*, 30: 201–214.

Lucas, P. J., de Bruijn, N. C., Schurink, K., & Hoepelman, A. (2000). A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in medicine*, 19: 251–279.

Naderpour, M., Lu J., & Zhang, G. (2014). An intelligent situation awareness support system for safety-critical environments. *Decision Support Systems*, 59: 325-340.

Pearl J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, CA: Morgan Kaufmann Publishers.

Pearl J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Qiu, J., Wang, Z., Ye, X., Liu, L., & Dong L. (2014). Modeling method of cascading crisis events based on merging Bayesian Network. *Decision Support Systems*, 62: 94-105.

Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *Knowledge Engineering Review*, 16 (3): 255–69.

Rice, M. E. & Harris, G. T. (1995). Violent recidivism: assessing predictive validity. *Journal of Consulting and Clinical Psychology* 63: 737–748.

Rice, M. E., & Harris, G. T. (2005). Comparing Effect Sizes in Follow-Up Studies: ROC Areas, Cohen's *d*, and *r*. *Law and Human Behavior*, 29: 5, 615-620.

Ronald, A., Mackoy, R., Thompson, V. B., & Harrell, G. (2004). A Bayesian Network Estimation of the Service-Profit Chain for Transport Service Satisfaction. *Decision Sciences*, 35 (4): 665-689.

Salini, S., & Kenett, R. S. (2009). Bayesian networks of customer survey satisfaction survey data. *Journal of Applied Statistics*, 36 (11): 1177-1189.

Sebastiani, P., & Ramoni, M. (2001). On the use of Bayesian networks to analyze survey data. *Research in Official Statistics*, 4 (1): 53-64.

Shaw, J., Doyle, M., Archer-Power, L., Wainwright, V., Hunter-Didrichsen, A., Dewa, L., Stevenson, R., and Forsyth, K. (2013). The Validation of New Risk Assessment Instruments for Use with Patients Discharged from Medium Secure Services (VoRAMSS).

Singh, J. P. (2013). Predictive Validity Performance Indicators in Violence Risk Assessment: A Methodological Primer. *Behavioral Sciences and the Law*, 31: 8-22.

van der Gaag, L. C., & Renooij, S. (2001). Analysing sensitivity data. In: *Proceedings of the 17th International Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, 530–7.

van der Gaag, L. C., Renooij, S., Witteman, C. L. M., Aleman, B., & Taal, B. G. (1999). How to elicit many probabilities. In: *Proceedings of the 15th International Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 647–54.

van der Gaag, L. C., Renooij, S., Witteman, C. L., Aleman, B. M., Taal, B. G. (2002). Probabilities for a probabilistic network: a case study in oesophageal cancer. *Artificial Intelligence in Medicine*, 25: 123–148.

Wang, G. A., Atabakhsh, H., Chen, H. (2011). A hierarchical Naïve Bayes model for approximate identity matching. *Decision Support Systems*, 51: 413-423.

Wu, P. P., Fookes, C., Pitchforth, J., & Mengersen, K. (2015). A framework for model integration and holistic modelling of socio-technical systems. *Decision Support Systems*, 71: 14-27.

Yet, B., Perkins, Z., Fenton, N., Tai, N. & Marsh, W. (2013a). Not just data: A method for improving prediction with knowledge. *Journal of biomedical informatics*, DOI: 10.1016/j.jbi.2013.10.012.

Yet, B., Bastan, K., Raharjo, H., Lifvergren, S., Marsh, W. & Bergman B. (2013b). Decision support system for Warfarin therapy management using Bayesian networks. *Decision Support Systems*, 55 (2), 488-498.

Yet, B., Constantinou, A., Fenton, N., Neil, M. Luedeling, E., & Shepherd, K. (2015). Project Cost, Benefit and Risk Analysis using Bayesian Networks. *Proceedings of the 12th UAI Bayesian Modeling Applications Workshop co-located with the 31st Conference on Uncertainty in Artificial Intelligence (UAI 2015)*, Amsterdam, Netherlands, July 12-16, 2015.