RC25671 (IRE1801-005) January 9, 2018
Computer Science

IBM Research Report

Vulnerability Models Allowing for Incomplete Information and Imprecision Application to Eldercare

Lea A. Deleris, Pol Mac Aonghusa, Robert Shorten*
IBM Research
IBM Dublin Technology Campus, Bldg. 3
Damastown Industrial Estate
Mulhuddart
Dublin 15, Ireland

*Also with University College, Dublin, Ireland

Vulnerability Models Allowing for Incomplete Information and Imprecision Application to Eldercare

Léa A. Deleris[a,*], Pol Mac Aonghusa[a], Robert Shorten[a,b]

[a] IBM Research Ireland, Dublin, Ireland

[b] University College Dublin, Dublin, Ireland

* Corresponding author: lea.deleris@ie.ibm.com

Abstract

Personalized, multi-dimensional assessments of the risk of the condition of a person (or object) deteriorating, called vulnerability assessments, are critical in outcome-based care management. However, lack of time for information gathering, often coupled with urgency to take action, pose specific challenges for individual vulnerability assessment in the field compared with assessment in a clinical or laboratory setting. We describe an approach to develop vulnerability assessment models for use "in the wild" - as distinct from use in clinical/laboratory settings - borrowing a term from the visual emotion recognition terminology. Key elements of the proposed modelling framework for vulnerability assessment are that it (i) accommodates incomplete information about the person (ii) continuously adjusts as background information evolves (iii) can serve as a guide to prioritize information gathering (iv) can function with imprecision in the input parameters. Specifically, we integrate a Markov Chain model describing the evolution of the person into, and out of, vulnerable states together with a Bayesian network that serves to customize the dynamic model. We describe the extension of the framework to situations arising through the presence of imprecision in the model parameters. The examples that we present were

developed and validated as practical tools for the analysis of the vulnerabilities of elderly persons in the context of a consulting engagement in China. The techniques presented are general however, and generalizations are discussed in the conclusion section.

Keywords: Vulnerability; Bayesian Networks; Markov Chains; Imprecise Probability; Health and Social Care; Ranking algorithms;

1. INTRODUCTION

Health and social programs represent a significant – and growing – proportion of the public sector budget across OECD economies, recently estimated at 22% of GDP [1]. In the US, Medicaid spending, $457 billion in 2014, is expected to almost double by 2020 [2]. As expenses continue to grow, it has become essential for sustainability to investigate solutions for reducing health and social care expenses while possibly minimizing the effect on the support provided (if not to improve its quality at the same time).

At an operational level, vulnerability models, i.e., models that estimate the risk of individual situations deteriorating, can help care workers identify the future high-cost / high-need patients and attempt to proactively address their problems. Vulnerability models constitute promising avenues to address the health and social care spending challenge. Indeed, social care services have typically been designed to address specific situations and tend to be ill-suited in more complex scenarios, i.e., when a person presents a variety of health and social needs that require coordination across social services. In addition, such complex situations are often associated with high costs. In NYS about 5% of Medicaid beneficiaries account for more than half of Medicaid spending [3] and in Camden NJ, one percent of patients account for a third of the city's medical costs [4].

In this paper, building upon previous work [5], we present such a vulnerability model for social care. Its purpose is to assist care coordinators (whether healthcare or social care) in assessing clients as they receive new situations or update existing ones. Our model provides them with a high level assessment of the patient's current needs, along with assessments of their vulnerabilities (upcoming problems). The objective is to help care workers with limited time and financial resources define priority domains for intervention for each individual and also compare profiles among individuals. Ultimately, the expectation is that the use of such a vulnerability model would lead to earlier identification and management of high-cost / high-need situations.

In our approach, vulnerability is measured through an estimation of how soon one would be in unwanted states, providing an indication of the urgency of the patients' situations along with the domains in which they are the most vulnerable. An essential characteristic of our model is that it adapts to the amount of information at hand, providing coarse estimates when few pieces of information about the person are available (as in the case for first visits) and updating vulnerability indexes as knowledge about the patient's context accumulates through interactions with the members of the care team. As such it is purposely designed for use in real world settings. As a side benefit, the model promotes shared use of information among care team members, encouraging better information exchanges of social and health services across silos.

2. RELATED WORK

As an introductory side note, we wish to clarify that what we term vulnerability model is different from social vulnerability indexes that also appear in the risk analysis literature [6]. Those social vulnerability indexes aim at evaluating which geographical zones are most vulnerable to hazards from a social perspective, accounting for likelihood of various types of disasters along with the effectiveness of recovery efforts. In our context, a vulnerability model typically applies to a person, not an area, and seeks to estimate the likelihood and severity of deterioration of that person situation along multiple dimensions.

We now focus on research linked to social care vulnerability models as we define them. The social care and healthcare literatures provide several examples of models designed for assessment or prediction of vulnerability, however with the objective of further population-level statistical analysis and based on complete gathering of information. By contrast, our objective is to develop models that function with limited and incomplete information and are aimed at decision support. We summarize nonetheless the main models that we have encountered in vulnerability prediction as they serve as informative guides to our modeling effort in eldercare (e.g., choice of explanatory variables). In the social care domain, InterRAI(7) and Northern Ireland Single Assessment Tool (8) are a suite of coordinated assessment tools for eldercare. Both tools are descriptive, seeking to capture and record an accurate profile of the patient at a given time. Other seek to estimate incidence rates of disabilities so as to predict the size of the elderly population that would require support in the coming years and inform public policy decisions (9). In the medical domain, frailty relates to the increased vulnerability of some elderly people to functional decline and dependence and encompasses biomedical, social and psychological aspects (10). More than twenty frailty instruments, questionnaires designed to measure frailty for use in clinical studies (such as population level trends), have been identified in the literature since the late 1990s (11,12,13).

Modeling-wise, risk analysis has long made use of Bayesian networks and Markov models. Bayesian networks, often represented in the less compact form of event trees, are one of the key models in probabilistic risk analysis and have for instance been applied to support risk management in health care entities (14), catastrophic risk modeling (15) also extended to counter terrorism (16). Markov chains, whether discrete or continuous, have also been widely used to represent the stochastic evolution of systems or incidents in particular in reliability modeling (17), but also to model adversarial situations (18). Sometimes Markov chains and Bayesian networks are combined, for instance in a fire modeling application (19). If we frame vulnerability modeling as a ranking task (ranking the different domains of vulnerability for an individual and across individuals), we note that Markov chains have also been used extensively, in particular through the well-known PageRank model [20] for ranking internet pages. They have since been applied to a variety of settings such as neurosciences, sports and literature [21]. However, there seem to be limited investigation for vulnerability modeling and applications in the social care domain.

Our technical contribution in this research goes beyond the application of Markov chains and their combination with Bayesian networks to a different context. Our approach addresses two key technical challenges associated with vulnerability assessment in the field – lack of comprehensive data, and, uncertainty in inputs. To compensate for lack of data, our approach *integrates* Bayesian networks and Markov models together, using the Bayesian network to enable continuous personalization of the Markov model. Specifically, the Bayesian network enables evaluation of vulnerability with varying amounts of information about the person. Through inference, limited evidence is leveraged to better estimate the parameters of the Markov model representing the dynamic vulnerability model of an individual. In addition, we explore how existing extensions of Bayesian network models and Markov Chains can be leveraged to accommodate imprecision in the model parameters. This is useful in practice to allow for limited knowledge about such parameters due to conflicting experts opinions or unavailability of large datasets for training.

3. PERSON-SPECIFIC VULNERABILITY MODELING FRAMEWORK

Our vulnerability modeling framework consists of a single integrated model that simultaneously evaluates a set of vulnerability metrics across several domains. In the context of eldercare, for instance, the vulnerability domains used were “Physical Health”, “Mental Health”, “Income”, “Shelter” and “Difficulty Walking”. For each domain, we compute a vulnerability metric corresponding to an estimation of how soon the person of interest might be in a state defined as “vulnerable” (for example, being homeless or being unable to walk independently). This can then be used to rank intervention.

3.1 Using a Markov Chain as Underlying Dynamic Model

The heart of a vulnerability model is a dynamic representation of the probabilistic evolution of an individual through states over time. We rely on a homogeneous Markov chain which is a random process characterized by the Markov property. This property, which stipulates that the evolution of the system only depends on its current state and not on how it reached that state, enables to characterize the model with few parameters and simplifies computations. As our goal is to consider multiple dimensions of vulnerability simultaneously, the Markov chain needs to consider state combinations, as in the example on Figure 1 which considers together Shelter and Employment vulnerabilities.

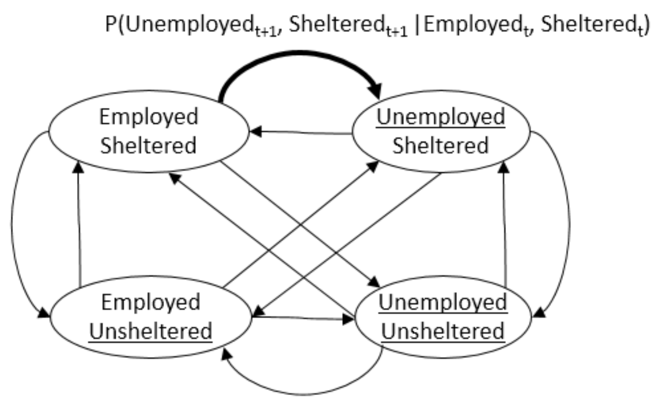


Figure 1– Simple Markov Chain Model to Represent Evolution of a Person over time

Note: To make the diagram readable, we have not represented the self-transitions. The transition probability that is spelled out is associated with the bolded arc.

Based on the Markov chain parameters, we compute future-behavior statistics aimed at characterizing the vulnerability, specifically, the mean first passage time (MFPT), defined shortly, to a vulnerable state conditional on the current state. In our example, assuming the person current state is "Employed, Sheltered", then the Employment vulnerability would be the average time of the first "visit" to either

"Unemployed, Sheltered" or "Unemployed, Unsheltered" states. MFPT provides a characterization of the short-term behavior of the model and is thus particularly suited as a measure of immediate vulnerability of a person. The use of mean first passage time as a mean to rank by decreasing level of relevance can be found in other domains, for instance in the context of road network modeling[22].

Mathematically, Let *d* index vulnerability domains and let D denote the number of such domains. Each domain is described by a state space $\boldsymbol{\mathcal{S}^d}$ decomposed into two subsets $\boldsymbol{\mathcal{S}^d_{vul}}$ and $\boldsymbol{\mathcal{S}^d_{ok}}$. The subset $\boldsymbol{\mathcal{S}^d_{vul}}$ corresponds to the states that are deemed vulnerable. In our preceding example, we would have D = 2. The first domain is "Shelter" with **state space** $\boldsymbol{\mathcal{S}^1} = \{\boldsymbol{Sheltered}, \boldsymbol{Unsheltered}\}$, **and vulnerable state set** $\boldsymbol{\mathcal{S}^1_{vul}} = \{\boldsymbol{Unsheltered}\}$, and the second dimension is **"**Employment**"**, with **state space** $\boldsymbol{\mathcal{S}^2} = \{\boldsymbol{Employed}, \boldsymbol{Unemployed}\}$, **and vulnerable state set** $\boldsymbol{\mathcal{S}^2_{vul}} = \{\boldsymbol{Unemployed}\}$. We denote by $\boldsymbol{\mathcal{S}}$ the product space of each state space for each dimension. $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}^1} \times \cdots \times \boldsymbol{\mathcal{S}^D}$. Let N denote the total number of states, $\boldsymbol{N} = \boldsymbol{card}(\boldsymbol{\mathcal{S}})$.

Let $\boldsymbol{x_t^d}$ represent the state of a person at discrete time intervals *t* along vulnerability domain *d*. Thus, $\boldsymbol{x_t} = (\boldsymbol{x_t^1}, \cdots, \boldsymbol{x_t^D})$ captures the comprehensive state of the person at time *t*. We assume that there exists a homogeneous Markov chain with transition matrix $\boldsymbol{M}$ with $\boldsymbol{m_{ij}} = \boldsymbol{P}(\boldsymbol{x_{t+1}} = \boldsymbol{s_j} | \boldsymbol{x_t} = \boldsymbol{s_i})$ where $\boldsymbol{s_i}, \boldsymbol{s_j} \in \boldsymbol{\mathcal{S}}$, which describes the evolution of a person across states.

Finally, let $\tau_{ij} = \min\{\, t : x_t = \boldsymbol{s_j}\}$ represent the random variable capturing the number of steps to arrive at destination $\boldsymbol{s_j}$ for the first time starting from state $\boldsymbol{s_i}$ at $t = 0$. The mean first passage time $\boldsymbol{\mu_{ij}} = E(\tau_{ij})$ is simply the expectation of this random variable and follows (1) below, where with probability $\boldsymbol{m_{ij}}$ it takes one step to go from state $\boldsymbol{s_i}$ to state $\boldsymbol{s_j}$ and for transition to states $\boldsymbol{s_k}$ with $k \neq j$, it would then take an expected time of $\boldsymbol{\mu_{kj}}$ to reach state $\boldsymbol{s_j}$ in addition to the current transition:

$$\mu_{ij} = m_{ij} + \sum_{\{k \neq j\}} (m_{ik} + 1)\, \mu_{kj} = 1 + \sum_{\{k \neq j\}} m_{ik}\, \mu_{kj} \qquad (1)$$

Building about this concept, for each domain d, we define the vulnerability index associated with current state $\boldsymbol{s_i}$ in a similar manner. We need to adapt the definition to the multidimensional setting and also to the fact that we want to evaluate the mean first passage time to a set of states (the ones tagged as vulnerable) rather than to a single state. In particular, let $\tau_i^d = \min\{\, t : x_t^d \in \boldsymbol{\mathcal{S}_{vul}^d}\}$ denote the number of steps to arrive to any destination state for which its dth component is deemed vulnerable starting from state $\boldsymbol{s_i}$. Then the vulnerability index $\boldsymbol{v_i^d}$ is simply the expectation of this random variable, $\boldsymbol{v_i^d} = E(\tau_{ij})$. When the starting state corresponds to a vulnerable state for domain d ($\boldsymbol{s_i}^{\boldsymbol{d}} \in \boldsymbol{\mathcal{S}_{vul}^d}$) our chosen convention is to set the vulnerability index to 0 to indicate current need. Finally, we call the vector $\boldsymbol{V_i} = (\boldsymbol{v_i^1}, \cdots, \boldsymbol{v_i^D})$, the vulnerability profile of the person corresponding to current state $\boldsymbol{s_i}$.

Slightly adapting equation (1) to our specific context, we have

$$v_i^d = 1 + \sum_{\{k: \boldsymbol{s_k}^{\boldsymbol{d}} \in \mathcal{S}_{ok}^d\}} m_{ik}\, v_k^d \quad \forall i: s_i{}^d \in \boldsymbol{\mathcal{S}_{ok}^d} \tag{2}$$

Let $M_{[ok]}^d$ denote the submatrix of M restricted to indexes i: $s_i{}^d \epsilon\ \boldsymbol{\mathcal{S}_{ok}^d}$, and similary $v_{[ok]}^d$ denote the vulnerability indexes vector for states $\forall i: s_i{}^d \epsilon\ \boldsymbol{\mathcal{S}_{ok}^d}$. Then, according to Equation (2), we have

$$(I - M_{[ok]}^d) v_{[ok]}^d = \mathbf{1}. \tag{3}$$

$M_{[ok]}^d$ is by design substochastic. In addition, we assume there are no absorbing states among the non-vulnerable states therefore all states in $M_{[ok]}^d$ are transient and $\lim_{n\to\infty} (M_{[ok]}^d)^n = 0$. Overall, $I - M_{[ok]}^d$ is non-singular and we have

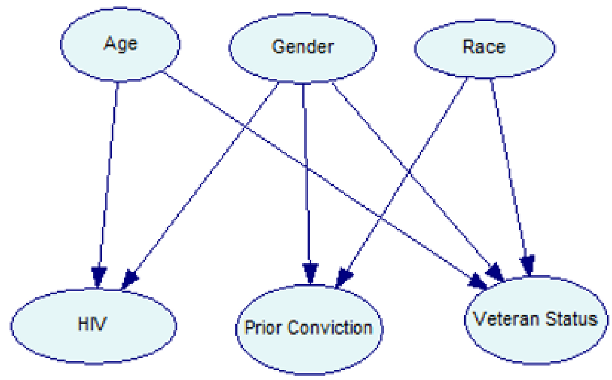
$$v_{[ok]}^d = (I - M_{[ok]}^d)^{-1} \mathbf{1}. \tag{4}$$

Overall, the vulnerability profile of a person is fully determined by the knowledge of the current state and the transition matrix $\boldsymbol{M}$. As we have pointed out, we may not know exactly what the current state of the person is. In that case, we will assume that we have some knowledge of the distribution of the current state of the person (as will become clearer when we introduce the Bayesian networks in section 3.2) which enables us to average over vulnerability indexes. Section 3.4 will provide some perspective on the challenge of having limited information about the transition matrix.

3.2 Bayesian Network for Customization

Observe that if the parameters of the Markov chain are fixed, any person in a similar current state would have the exact same vulnerability profile, providing minimal differentiating information. To allow for model customization, we have chosen to let the parameters of the Markov chain depend on a set of relevant factors. Specifically, we assume that there exists an underlying Bayesian network [(23)] which articulates the relationships among the factors themselves and also between the factors and the state of a person. Figure 2 shows an example of such a Bayesian network with six factors. This model implies that *Age*, *Gender* and *Race* are independent of one another but that *Veteran Status* is influenced by all three. In addition to the graphical structure, Bayesian networks are characterized by a quantitative layer which represents the distributions of the variables conditional on all possible parent scenarios (See for instance at the bottom of Figure 2, the conditional probability table for the variable *HIV positive)*. Bayesian networks are especially efficient at computing inference queries, i.e., conditional probabilities involving the subsets of the variables in the network. In our example such a query could be asking for the probability of a person being *HIV positive* given *Gender* and *Veteran Status*.

Note that we can extend such a model to include variables corresponding to a Markov chain model as presented in Figure 3. By extending the Bayesian network so as to couple it with the Markov chain model, inference computations can provide personalized parameters for the Markov chain, accounting for the current knowledge of a person situation, both current state and known factors, as large or limited as this knowledge may be.



		Male Age:20-50	Male Age:50+	Female Age:20-50	Female Age:50+
HIV Positive	Yes	p_1	p_2	p_3	p_4
	No	$1\text{-}p_1$	$1\text{-}p_2$	$1\text{-}p_3$	$1\text{-}p_4$

Figure 2– Illustrative Bayesian Network

Mathematically, Let $\boldsymbol{\chi} = \{\boldsymbol{X_1}, \cdots, \boldsymbol{X_n}\}$ denote the set of all the variables represented in the Bayesian network. We decompose $\boldsymbol{\chi}$ into three groups:

- $\mathcal{F} = \{ x_k^F : k = 1, \cdots, K\}$: set of variables representing explanatory factors
- $\mathcal{S}_t = \{ x_t^d : d = 1, \dots, D\}$: set of variables representing the state variables at current time t
- $\mathcal{S}_{t+1} = \{ x_{t+1}^d : d = 1, \dots, D\}$: set of variables representing the state variables at future time t+1

The last two sets enable the pairing. Indeed, each element in the set $\boldsymbol{S_T}$ corresponds to one of the vulnerability domains and $\boldsymbol{Val}(\boldsymbol{x_t^d}) = \boldsymbol{Val}(\boldsymbol{x_{t+1}^d}) = \boldsymbol{\mathcal{S}^d}$. Where in a Markov model, a state *s* was represented by one (vector) variable, it is represented by a set of variables in the Bayesian network. For any observed evidence $\boldsymbol{\mathcal{E}}$ belonging to a subset of the state space of $\boldsymbol{\mathcal{F}}$, and any $(\boldsymbol{s_i}, \boldsymbol{s_j}) \in \boldsymbol{\mathcal{S}^2}$, the

Bayesian network model can provide through inference the value of the parameters of the Markov chain model :

$$P\left(x_{t+1}^{1}=s_{j}^{1},\ldots,x_{t+1}^{d}=s_{j}^{d},\ldots,x_{t+1}^{D}=s_{j}^{D}\middle|x_{t}=s_{i},\mathcal{E}\right) \quad (5)$$

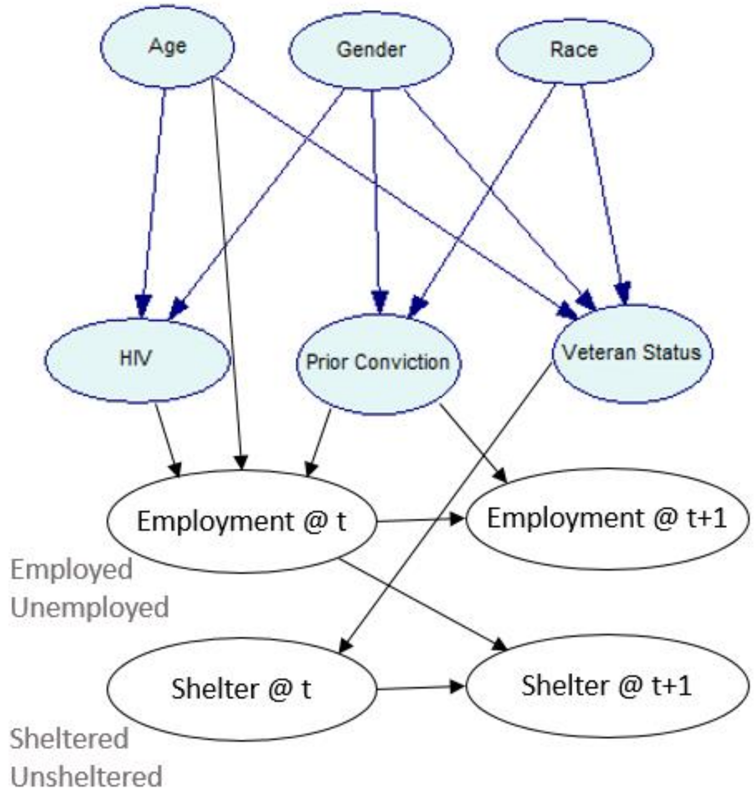


Figure 3– Illustrative Paired Bayesian Network

Therefore, when we let $\mathcal{E}$ represent the observed information about a person related to factors in $\mathcal{F}$, we can estimate the associated personalized Markov chain parameters and thus that person's vulnerability profile. Consequently, as knowledge about a person's situation evolves, the vulnerability indexes evolve as well. In addition, the Bayesian network enables us to estimate $P\left(x_{t}^{d}\middle|\mathcal{E}\right)$ i.e., the distribution of the current state of person along domain *d* when it is not part of the observed evidence. This enables us, as noted as the end of the previous section, to derive a vulnerability index even when ignoring the current state of a person.

3.3 Characteristics of the Modeling Framework

This modeling framework presents several advantages. First, the model is designed to function with *incomplete information* about a person current state and profile, a key difference for instance with assessment tools in the social care domain, making it a useful complement for screening and other decision making situations "in the wild'. Second, through the Bayesian network model, the vulnerabilities across domains are made interdependent. In the context of health and social care, this means that the model can capture essential *interactions* among social factors, health factors and vulnerabilities. Third, the resulting vulnerability profile is both *personalized yet standardized*. It is personalized as the Bayesian network provides a way to determine the parameters of the Markov chain model based on observed state of the person. It is standardized as the same customization model is used for everyone. Note also that the vulnerability indexes are associated with a *physical – thus interpretable* – measure, in the form of average amount of time. This consistent use of one metric enables both within subject and across subject comparisons.

In the presentation of the model above, we have focused on measuring vulnerability through the risk of occurrence of a vulnerable state. The model can be used at the same time to evaluate severity upon occurrence. Specifically, for a vulnerability domain *d*, for a state $\boldsymbol{s_i}^{\boldsymbol{d}}$ corresponding to a vulnerable state, then the severity index $\boldsymbol{w_i}^{\boldsymbol{d}}$ is defined as the mean first passage time to a non-vulnerable state $\left\{\boldsymbol{s} \in \boldsymbol{\mathcal{S}}: \boldsymbol{s^d} \notin \boldsymbol{\mathcal{S}^d_{vul}}\right\}$. For a state $\boldsymbol{s_i}^{\boldsymbol{d}}$ that does not correspond to a vulnerable state, then we can set $\boldsymbol{w_i}^{\boldsymbol{d}} = \boldsymbol{0}$, symmetrically to the definition of severity.

Besides the vulnerability profile, the model yields further insights into a person's situation. In particular, we can provide (i) guidance into what additional information could be useful (informative factors) and (ii) explanations (influential factors) by identifying which aspects of the person profile strengthen and weaken his/her profile. Both are obtained by performing one-way sensitivity analyses. For instance, let's consider domain d, and current vulnerability index $\boldsymbol{v}_i^d(\boldsymbol{\mathcal{E}})$, corresponding to knowledge $\boldsymbol{\mathcal{E}}$. For each factor $\boldsymbol{x}_k^F$ currently observed ($\boldsymbol{x}_k^F \in \boldsymbol{\mathcal{E}}$), we compute an artificial vulnerability index $\widetilde{\boldsymbol{v}}_i^d(\boldsymbol{\mathcal{E}} \backslash \{\boldsymbol{x}_k^F\})$, assuming factor k is not observed. The difference between the vulnerability indexes, $\boldsymbol{v}_i^d(\boldsymbol{\mathcal{E}}) - \widetilde{\boldsymbol{v}}_i^d(\boldsymbol{\mathcal{E}} \backslash \{\boldsymbol{x}_k^F\})$, provides an indication of the influence of that factor. A similar prodecure is performed for factors that are unknown to determine how informative they could be, loosely borrowing from the concept of Value of Information from the field of Decision Analysis.

3.4 Extension to Imprecision: Credal Networks and Markov Sets

One useful characteristic of the approach that is relevant when building models from real world information is the fact that it can function with imprecise information. So far, we have assumed that the parameters of the model are point estimates, i.e., that we have enough data or definite domain expertise to determine a single value for each one. This may not necessarily be the case in practice, either because data is limited, experts are unsure or more comfortable with providing ranges, or because multiple experts have proposed different values. In such situations, being able to handle imprecise information is essential if one does not have the time or budget to reconcile conflicting information or gather more data.

Both Markov chains and Bayesian networks admit imprecise extensions in the form of Markov Set Chain [24] and Credal Network [25] (respectively) which enable to rigorously percolate the imprecision through the computations. The only consequence of allowing for imprecision in the model is the fact that vulnerability indexes become intervals and may thus be more difficult to interpret.

Specifically, assume we are asking about the effect of X on Y, experts may not be able to provide point estimate of P(Y|X) but may provide a range $[\underline{p}, \overline{p}]$. Such type of imprecision can be incorporated in the model both at the Bayesian network level and at the Markov chain level.

For Bayesian networks, imprecision in the form of ranges can be modeled through a credal network. A credal network is an extension of a Bayesian network where instead of associating conditional probability distributions to the nodes, one associates conditional credal sets. A credal set $K(X|Pa(X) = \pi_i)$ for all possible instantiations π_i of the parents of node *X* denoted $Pa(X)$, is a closed convex set of probability distributions. For instance if X is binary taking values x and $\neg x$, then $K(X|Pa(X) = \pi_i)$ can be represented by a closed interval and denoted $[p,q] = \{P(X): p \leq P(X = x) \leq q, P(X = x) + P(X = \neg x) = 1\}$. Rules for combining credal sets and defining independence that extend Bayesian networks manipulation are to be specified [25]**.** One choice for instance is to use strong extension for combining credal sets which leads to independence properties quite similar to standard theory of Bayesian networks[25]. Naturally, the representation being more sophisticated, all operations such as inference and expectation require significantly more memory and computing power that the precise Bayesian network framework. For our purpose, we simply focus on the fact that from a credal network, it is possible to obtain coherent lower (denoted $\underline{p}(X|Y)$) and upper probabilities (denoted $\overline{p}(X|Y)$) for any conditional event *X*/*Y* by deriving the credal set K(X, Y) and defining $\underline{p}(X|Y) = inf_{p \in K(X,Y)} p(X|Y)$, $\overline{p}(X|Y) = 1 - \underline{p}(X^c|Y)$, where $X^c U X = \Omega$.

In our specific case, it means that from the credal networks we are able to extract intervals for the transition probabilities in the Markov Chain. Fortunately, the theory of Markov chains has also been extended to handle imprecision in the form of ranges through models called Markov Set-Chains [24]. In such model, the transition matrix A is replaced by compact set of $n \times n$ stochastic matrices denoted M. We define $M^{k+1} = M \times M^k$ as the set of matrices being the product of *k+1* matrices belonging to the set M, thus $M^{k+1} = \{A_1 \cdots A_{k+1}: A_1, \cdots, A_{k+1} \in M\}$. The sequence$M$, M^2,… is called a Markov set-chain. Specifically, if we look at imprecision defined through interval, then we have the following:

Let p and q be $1 \times n$ vectors with $p \leq q$ (component-wise), then we define a vector interval

$[p, q] = \{x: x \; is \; a \; 1 \; x \; n \; stochastic \; vector \; and \; p \leq x \leq q\}$.

If $p_i = \min_{x \in [p,q]} x_i$ and $q_i = \max_{x \in [p,q]} x_i$ then p_i and q_i are called tight respectively. If p_i and q_i are tight for all *i* then the interval $[p, q]$ is called tight. Similarly, a matrix interval is defined by P, Q, two *n* x *n* non negative matrices with $P \leq Q$ (component-wise) and

$[P, Q] = \{A: A_i \in [P_i, Q_i] \; \forall i \; where \; A_i, P_i, Q_i, are \; the \; ith \; rows \; of \; A, P, Q \; respectively\}$

Similarly, if P and Q satisfy, $p_{ij} = \min_{A \in [P,Q]} a_{ij}$ and $q_{ij} = \max_{A \in [P,Q]} a_{ij}$ for all *i* and all *j* then $[P, Q]$ is called tight. In particular, if $M = [P, Q]$ is a column-tight interval for a transition matrix and $[L, H]$ are tight component bounds on any M^k then it is possible through the Hi-Lo algorithm to determine $[\overline{L}, \overline{H}]$ corresponding to tight components for M^{k+1} [24]. Furthermore, the Hi-Lo method is also involved in computing bounds for the mean first passage time. Note that by definition, the bounds generated by lower and upper probabilities are tight.

Altogether, the Bayesian network extended to a credal network results in being able to define a Markov-set-chain $[P, Q]$ associated with any evidence about a person which in turn can be used to compute lower and upper bound on vulnerability indexes.

4. APPLICATION TO ELDERCARE

We applied this modeling framework to the domain of eldercare in collaboration with Beijing Academy of Science and Technology. Specifically, we developed a vulnerability model to help care workers better support aging populations, focusing on fostering independent living while at the same time identifying declining mental and physical health as early as possible. Aging populations and support for independent living are a growing concern in both developed and developing countries. In fact, the number of older people who are no longer able to look after themselves in developing countries may quadruple by 2050[26].

To build the Bayesian network, we have used the Longitudinal Study On Aging (LSOA)[1]from the US National Institute for Health. This data is based on an initial cohort of 9,447 persons who were 70 aged and over in 1995. Those persons (or their proxies when deceased) have been interviewed at three different times (waves), approximately two years apart. We have chosen to use the last two waves as the questionnaires were more similar between those two waves than between the initial and second waves. We thus obtained access to state variables at two-year intervals. The choice of vulnerability domains and higher level vulnerability categories was mostly driven by the availability of data and by local expert knowledge[27]. Table 1 provides a list of the vulnerability domains considered grouped into 6 categories. The LSOA study contains several hundred different fields and we have selected a subset of 61 features in addition to the 34 state variables (listed on the left side of the table). Compared with medical domain frailty assessments[28,29], our set of explanatory factors contains more social environment information (family members, education, type of household) and less focus on laboratory measurements (neurological exam, glucose or albumin measurements, grip strength). However, there is overlap in terms of Activity of Daily Living and Instrumental Activity of Daily Living along with indications of chronic conditions such as diabetes and hypertension.

After basic data cleaning (mostly discretizations of some fields and minimal data imputation), we obtained a training data set with 3604 entries after setting aside some entries for testing. The Bayesian network model was learnt using Jsmile[2] greedy-thick-thinning algorithm with various combinations of the learning algorithm parameters (type of priors and maximum number of parents allowed). While we do not have

[1] Description and data available at http://www.cdc.gov/nchs/lsoa/lsoa2.htm

[2] available at http://genie.sis.pitt.edu/

data to validate the value of the vulnerability indexes, we chose the best set of parameters for the learning algorithm by taking the one that had the best next-period prediction accuracy, which we computed by applying the Bayesian network prediction to test data that we had held from the training data.

Table 1 – List of Factors and Vulnerability Domains Included in the Eldercare Vulnerability Model

Factors	**Category(number of variables):** list of vulnerability domains
Difficulty Stooping, Difficulty Preparing Own Meal, Has Arthritis, Difficulty Controlling Urination, Difficulty Finger Grasping, Difficulty Walking Quarter Mile, Difficulty Walking 10 Steps, Difficulty Grocery Shopping, Has Diabetes, Number of Persons in Household, Marital Status, Length Married in Years, Age, Veteran Status, Gender, Gender of Second Person in Household Number Days Left House in Past 2 Weeks, Education, Race, Has Cancer, Bed , Bath and Kitchen on the same Floor, Use Stairs to enter Home, Proximity Children in Hours, Frequency See Children, Frequency Talk to Children, Number of Living Daughters, Number of Living Sons, Number of Living Sisters, Number of Living Brothers, Has Ever Worked, Height, Weight, Has Asthma, Has Bronchitis, Smoker, Ever Had Broken Hip, Has Osteoporosis, Number of Years of Osteoporosis, Difficulty Reaching Over Head, Difficulty Reaching Out, Has Hypertension, Ever Had Hypertension, Had Stroke, Ever Had Diabetes, Number Year Diabetes, Number Year Arthritis, Ever Had Asthma, Ever Had Bronchitis Emphysema, Number Year Bronchitis Emphysema, Number Year Asthma, Number Year Hypertension, Had Heart Disease, Number Year Heart Disease, Number Year Stroke, Ever Had Cancer, Is Social Activity Sufficient, Difficulty Doing Light Housework, Difficulty With Telephone, Difficulty Managing Money, Difficulty Managing Medication.	**Sensation (2):** Sight – Hearing **Social Involvement (2):** Socially Active with friends - Socially Active with relatives **Mental Health (2):** Frequency of sadness or depression, self-rated memory **Fundamental Daily Activity (6):** Ability to eat, bath, dress, walk , get out of bed or chairs and use the restrooms independently **Extended Daily Activities (2) :** Social Living Abilities, Cognitive Living Abilities **Medical (3) :** Self-rated Health, Injured from Fall, More than three chronic conditions

The model was then integrated and deployed into a real tool currently being deployed in a Beijing neighborhood. Figure 4 provides a snapshot of the user interface associated with a previous iteration of our model (with fewer categories). In that interface, we do not directly present the full vulnerability profile (meaning values for each of the 14 vulnerability domains) but rather aggregate each category by reporting the worst vulnerability index associated with that category (second column). In addition, we provide associated historical values (third column) and contributing factors (fourth column) to the care worker using the dashboard. While we chose to take the worst vulnerability index among indexes within the same vulnerability category, other possible approaches could have been to average across the indexes and list the union of informative and influential factors in the fourth column.

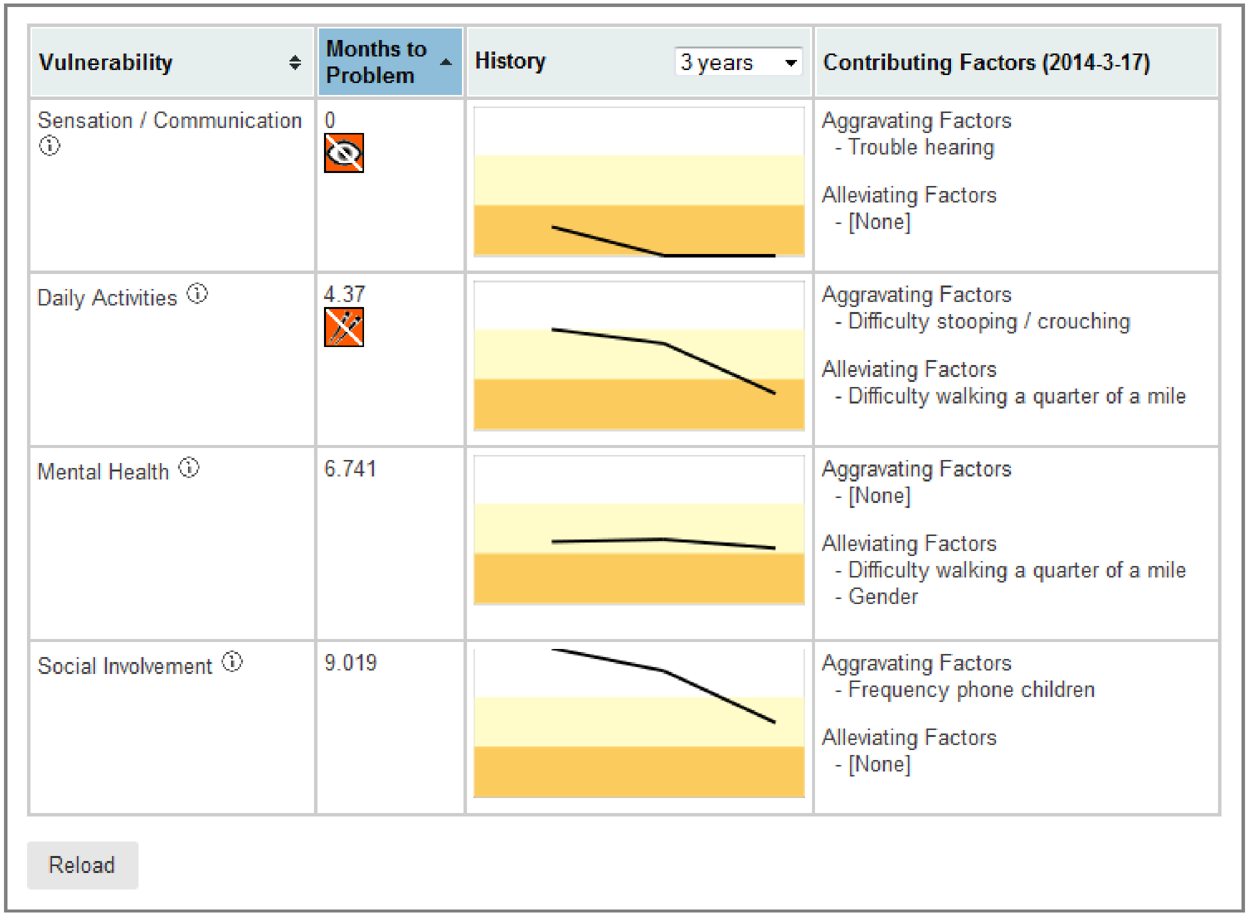


Figure 4– Dashboard of our Application

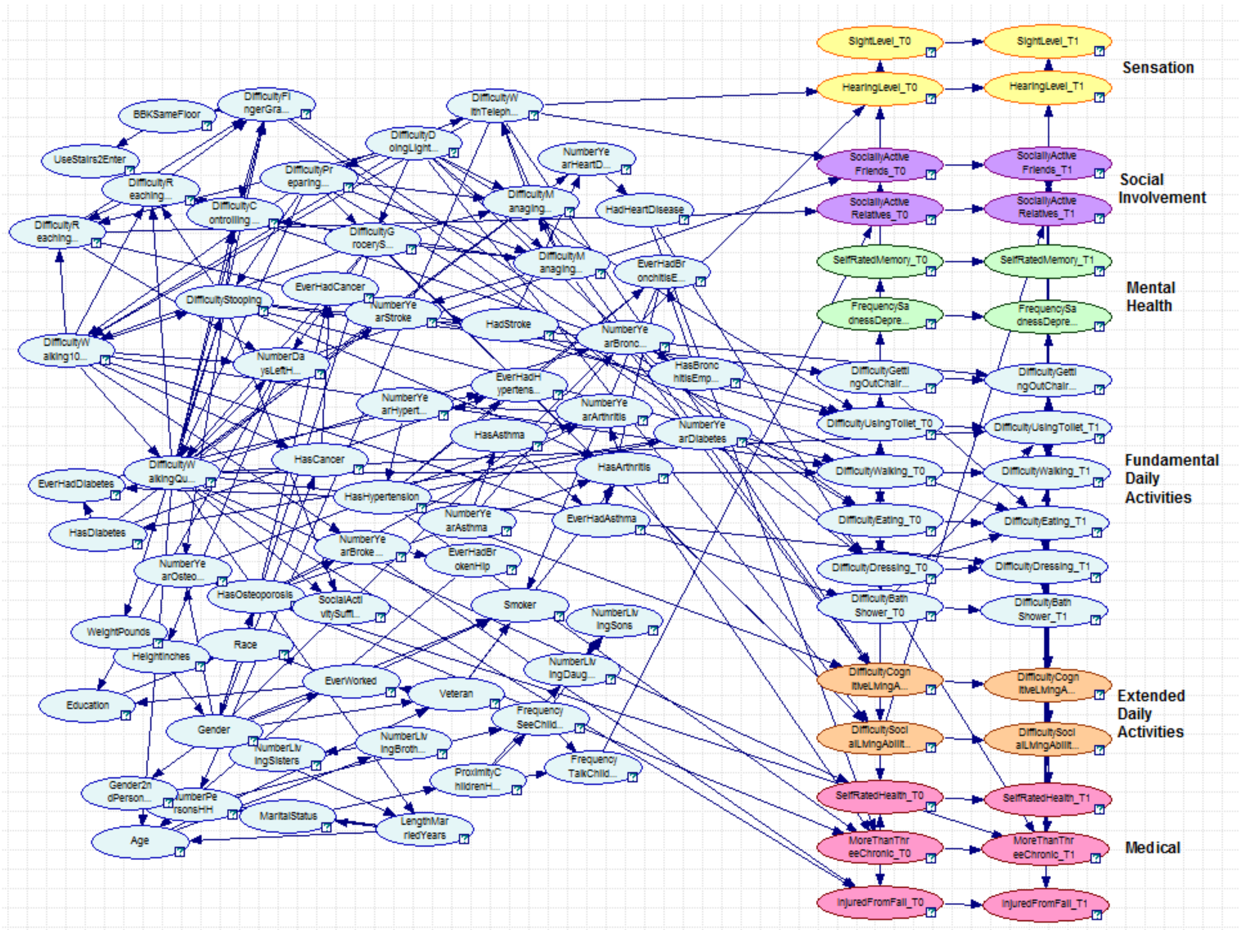


Figure 5– Bayesian Belief Network associated with our Model

To illustrate the effect of the known information on the vulnerability profile, we provide on Table 2 the profiles associated with 4 different information content associated with a patient. Person 1 corresponds to a "Married 75 year-old man". Person 2 corresponds to the same profile, except that we also know that the person suffers from diabetes. Person 3 builds upon person 2 with the additional knowledge that he has difficulty walking and feels depressed a little of the time. Finally person 4 is an even richer (and dire) description of the person current situation. However, person 4's profile still represents incomplete knowledge of the person, capturing 10 out of the 61 factors contained in the model.

The first two columns in Table 2 are fairly similar, though there is a consistent minimal increase of vulnerability (decrease of the indexes) for person 2 who also suffers from diabetes. This is an example of cross-influences of a medical factor on social domains. The largest difference is associated with the *More Than Three Chronic Conditions* domain. As diabetes is a chronic condition, such change is logical. When comparing persons 2 and 3, we first notice that *Difficulty Walking* is now null for person 3, a direct consequence from the fact that we have observed this difficulty, it is no longer a vulnerability but rather a need. Notice again that many other vulnerability indexes have decreased as well, illustrating cross-influences. Only *Frequency Sadness and Depression* has increased, indicating lessened vulnerability. This is not surprising as feeling sad or depressed only *a little of the time* constitutes in fact positive evidence compared to the unknown state. Finally person 4 represents a significant aggravation of person 3 state, with many additional physical and medical difficulties and a worsening of the mood.

In addition to providing vulnerability indexes, we also perform sensitivity analyses on the explanatory factors so as to determine which ones have a sizeable influence on the vulnerability index. For instance, for person 4 and for the domain of being socially active with friends, the index value is 9.29 years which seems to indicate little concern. However, our sensitivity analysis on known facts points out that the fact that person 4 has difficulty doing light housework and that he has left home less than 5 times in the past two weeks are aggravating factors which should be kept in perspective. By contrast, the sensitivity analysis on unknowns factors reveals that knowing whether person 4 has difficulty with using the phone, preparing his own meals, managing money or medication and difficulty doing grocery shopping are informative factors. This means that depending on the answer, the vulnerability index could change significantly.

Table 2 – Illustrative Vulnerability Indexes (in Years)

	Person 1	**Person 2**	**Person 3**	**Person 4**
Difficulty Dressing	7.51	7.49	4.42	1.12
Difficulty Eating	15.55	15.52	11.10	5.34
Difficulty Walking	3.24	3.22	***0.00***	***0.00***
Difficulty Using Toilet	8.83	8.81	5.11	2.14
Difficulty Getting Out Chair Bed	5.77	5.74	2.97	1.10
Sight Level	6.54	6.54	5.82	4.70
Hearing Level	2.70	2.66	2.58	2.26
Socially Active Friends	12.28	12.27	11.33	9.29
Socially Active Relatives	23.12	23.11	20.58	15.98
Frequency Sadness Depression	6.65	6.61	7.96	***0.00***
Self-Rated Memory	32.10	32.06	31.69	31.01
Difficulty with Social Living Abilities	4.05	4.04	2.59	0.43
Difficulty with Cognitive Living Abilities	4.30	4.29	3.37	1.55
Injured From Fall	4.45	4.45	3.56	2.83
More Than Three Chronic	9.10	***7.11***	6.13	3.79
Self-Rated Health	13.92	13.78	9.69	4.50

LEGEND:

Person 1: Married 75 year-old man

Person 2: Married 75 year old man with diabetes

Person 3: Married 75 year old man with diabetes and difficulty walking and feeling sad or depressed a little of the time

Person 4: Married 75 year old man with diabetes, hypertension, difficulty walking, stooping and finger grasping, feeling sad or depressed all of the time, who has not left his house more than 5 days in the past 2 weeks, who is a smoker and who is not able to do light house work in his home

Finally, we also evaluated severity indexes for the situations where vulnerability is no longer a risk but a need (i.e., when the index is 0). For person 3, the severity of difficulty walking is 5.4 years increasing to about 7.5 years for person 4, thereby serving as a further indicator that the condition of the person has worsened. For person 4, the severity of depression is 3.5 years highlighting here that addressing the difficulties with mobility may be the first priority for a care worker in charge of this person.

To illustrate the effect of imprecision on the results, we have reproduced a similar analysis though on a reduced set of variables and vulnerability domains as listed in Table 3.

Table 3 – List of Factors and Vulnerability Domains Included in the Eldercare Vulnerability Model with Imprecise Inputs

Factors	**Category(number of variables):**list of vulnerability domains
Difficulty Stooping, Difficulty Finger Grasping, Has Diabetes, Marital Status, Age, Gender, Number Days Left House in Past 2 Weeks, Has Ever Worked, Smoker, Ever Had Hypertension, Difficulty Doing Light Housework	**Mental Health** Frequency of sadness or depression,
	Fundamental Daily Activity Ability to walk independently

A precise model was learnt following the same protocol as with the broader set of variables. To incorporate imprecision, we modified some of the network parameters based on the number of data points in the training dataset that went into estimating the probabilities. Indeed, while we had about 3600 training entries, some specific combinations were very rare or did not occur altogether. For instance, no respondent above 90 was separated while in our network, the distribution of age is influenced by marital status. This makes the estimation of the probability of being above 90 given one is separated much less reliable than for married respondents (74 entries) or widowed ones (273 entries). Therefore, for parameters with a support of less than 5 entries, we replaced the precise probability by an interval corresponding to +/- 50% of the precise value. When parameters had a support above 5 but less than 20, we associated them with an interval +/- 10%. Granted, this represents a fairly simple approach to adding imprecision and the choice of the magnitude parameters or thresholds is arbitrary. Its purpose here is to provide some illustration as to cases where adding imprecision is warranted along with its effect on vulnerability modeling. To perform the credal set computations, we used the JavaBayes package[3].

Figure 6 reports the lower bound, upper bound of the mean first passage time to the vulnerable states for Difficulty Walking and for Depression starting from each of the possible non-vulnerable states. In addition, we also provide the values associated with the model without imprecision (red dots). We consider three profiles with increasing level of information about the persons (and with increased difficulties overall). The third profile corresponds to a 75 year old male with diabetes and difficulty stooping, finger grasping, doing light housework, has hypertension, currently smokes and had not left his house more than 5 days in the past two weeks. Before going into the specific results, we observe that this reduced model leads to higher vulnerability indexes than the broader model, around 5.5 years for difficulty walking for the first profile corresponding exactly to Person 1 in the broader model which had a vulnerability of 3.2 years. We observe that the range of the intervals varies with the level of knowledge, for the 75 year-old male for difficulty walking, we have a difference of 0.16 years between the lower bound and the upper bound while this difference becomes 0.40 years for the person with the many difficulties. This pattern applies to all three mean first passage time metrics. We also observe that the addition of imprecision does not necessarily lead to a symmetric interval around the reference value (the value corresponding to the model without imprecision). There is no reason why the interval should be symmetric as shown in our simple model. Finally, while in this illustrative example intervals tend not to overlap, we can see that for the case of becoming depressed from feeling sad a little of the time, the 75 year-old male with and without diabetes are difficult to distinguish and should probably be treated in a similar manner by a care worker.

[3] Available at http://www.cs.cmu.edu/~javabayes/Home/

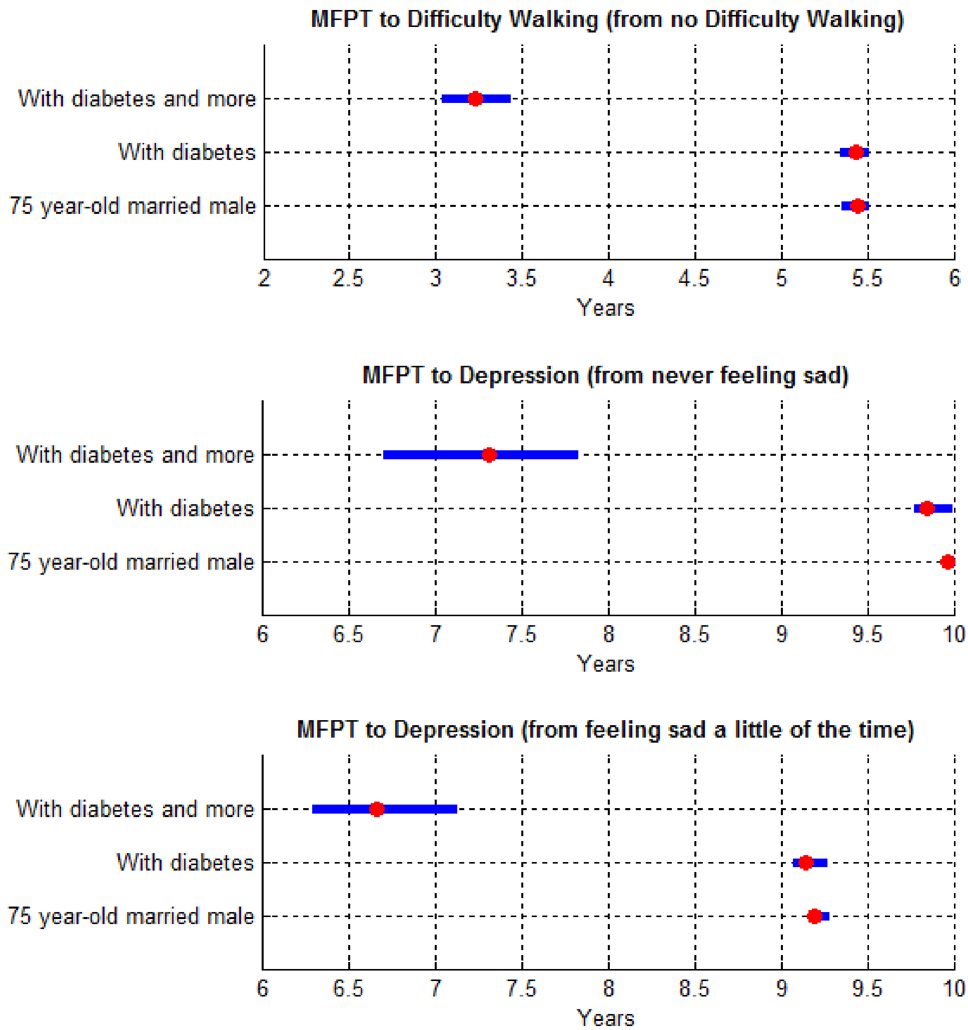


Figure 6 –Graphical Representation of Lower Bound, Upper Bound, and Reference Values for Multiple Vulnerability Domains Based on the Model With Imprecise Inputs

5. DISCUSSION AND CONCLUSION

In this paper, we present a modeling framework for vulnerability assessment “in the wild” and its application to the health and social care domain. This model combines a Markov chain model that describes the evolution of persons through states with a Bayesian network that enables to customize the Markov chain parameters based on available evidence. The objective of the model is to estimate the vulnerability profile of a person with varying amount of input information so as to guide subsequent decision making. Such a system is meant as a decision support tool to help professionals with limited time and financial resources to prioritize urgent situations and understand, for a given patient, which domain of vulnerability should be addressed first. Within the social care context, its purpose is among others to identify clients that present a multiplicity of problems and need to be handled differently than the majority of clients (the often termed “High-Cost High Need” category). Early identification of complex cases are expected to have positive benefits on both the persons involved and the efficiency of the social care system at large.

Our choice of underlying models, Bayesian networks and Markov chains, was partially motivated by the fact that they are fairly widely used and at the same time graphically intuitive. Altogether, this makes them accessible to decision makers that are not always analytically savvy yet interested in getting some understanding into logic of the model and the provenance of the vulnerability estimations.

One challenge associated with the presented modeling approach is to find a reasonable method to build the underlying Bayesian network (both structure and parameters). One technical advantage of our framework is its ability to incorporate imprecision in the parameters in a disciplined fashion through the combinations of Markov sets and credal networks. The ability to handle imprecise inputs makes the approach more likely to be applicable in real world situations where information can be messy. In the eldercare case study that we presented, we were satisfied with using the LSOA dataset to determine both model structure and parameters as it was reasonably large. However, further data sources could have been considered, for instance China Health and Retirement Longitudinal Study (CHARLS) [4], North American open data sets such as RAND Health and Retirement Study (HRS)[5] or similar studies from Europe such as SHARE[6], ELSA[7], or TILDA[8] among others. In fact as the model was designed to be deployed for a Chinese cohort, we would have preferred to use the CHARLS data, but it contained insufficient relevant entries when filtered to have respondents appearing in at least two waves. We decided, as a first approximation, to use the information from the American cohort. This meant not including potentially relevant cultural factors and assuming that aging factors were somewhat universal.

Overall, model building in this case, as in many other situations in risk analysis, is faced with the double challenge of *incomplete* or *partially misaligned* single input, which when one seeks to combine them may become *redundant* and *conflicting*. To facilitate the construction of the model, we are currently exploring methods to build Bayesian network models from a collection of disparate, overlapping, and possibly conflicting, sources of information such as

- Data sets at the individual level, covering a subset of the nodes in the network
- Aggregated information for a subset of the nodes of the network
- Expert information (influence statements and probability statements).

Specifically, we are extending methods from the artificial intelligence community. In particular, we have looked at adapting the PC algorithm[(30)] to build a Bayesian network from multiple overlapping datasets [(31)].

---

[4] Description available at http://charls.ccer.edu.cn/en
[5] Description available at http://www.rand.org/labor/aging/dataprod/hrs-data.html
[6] Description available at http://www.share-project.org/
[7] Description available at http://www.elsa-project.ac.uk/
[8] Description available at http://tilda.tcd.ie/

In a different endeavor focused on the medical academic literature, we have also explored how to extract expert statements from abstracts and then aggregate them into a Bayesian network [32].

Beyond eldercare and evident extensions, our vulnerability model can apply to many related situations that involve evaluating the risk of a person state deteriorating. One such domain would be in social actions for crime prevention, where it is critical to better understand who among the relatively large population of known perpetrators of crimes, or in a more particular case, of violent crimes (especially those who have committed lesser offenses such as shoplifting and possession of drugs), is likely to become a repeat offender. Again, to make such information actionable at an operational level, it needs to be personalized to each such candidate, taking into account the social-family context of the person along with information about the conditions in which the crime was committed. In particular, it would seem beneficial to be able to identify people who are likely to commit a crime in the near future. Another domain of application would be the management and maintenance of pipe networks, be they water networks as in cities or gas networks in hospitals. Seeing each pipe as a member of an aging population, the model would be able to estimate which ones are most likely to deteriorate based on contextual information about the pipe (for instance diameter, length, water pressure, material, location) and possible observation of leaks. Such a model would be useful for administrators to better manage the maintenance activities associated with the water network. However, as pipes in a network are not necessarily independent, this would require an extension of the model to allow for local interactions among the elements whose vulnerabilities are evaluated. This would represent in fact a meaningful extension even for the case of eldercare so as to be able to consider the cross-influences of a spouse and of a person's close social network on their well-being.

REFERENCES

1. OECD. Social spending after the crisis. 2012. OECD.

2. Centers for Medicare and Medicaid Services. Actuarial report on the financial outlook for Medicaid. 2012. *Washington, DC, US Government Printing Office.*

3. Mann C. Medicaid and CHIP: On the road to reform [pdf slides]. 2011. Retrieved from http://www.healthandwelfare.idaho.gov/Portals/0/Medical/SUD/Medicaid%20and%20CHIP%20On%20the%20Road%20to%20Reform%20by%20Cindy%20Mann.pdf.

4. Gawande A. The hot spotters. Can we lower medical costs by giving the neediest patients better care? The New Yorker. 2011. Retrieved from http://www.newyorker.com/magazine/2011/01/24/the-hot-spotters.

5. Deleris LA, Aonghusa PM, Shorten R. Person-Specific Standardized Vulnerability Assessment in Health and Social Care. MEDINFO 2015: EHealth-enabled Health: Proceedings of the 15th World Congress on Health and Biomedical Informatics 2015. 216:462-466

6. Cutter SL, Boruff BJ, Shirley WL. 2003. Social vulnerability to environmental hazards*. Social science quarterly. 84(2):242-61.

7. Gray LC, Berg K, Fries BE, Henrard JC, Hirdes JP, Steel K, Morris JN. Sharing clinical information across care settings: the birth of an integrated assessment system. BMC Health Services Research. 2009. *9*(1):71.

8. Taylor BJ. (2012). Developing an integrated assessment tool for the health and social care of older people. British journal of social work, 2012. *42*(7):1293-1314.

9. Van der Gaag N, Bijwaard G, de Beer J, Bonneux L. 2015. A multistate model to project elderly disability in case of limited data. Demographic Research. 32:75

10. Lang PO, Michel JP, Zekry D. Frailty Syndrome: A Transitional State in a Dynamic Process. Gerontology 2009. 55:539-549

11. De Vries NM, Staal JB, Van Ravensberg CD, Hobbelen JS, Rikkert MOM, Nijhuis-Van der Sanden MWG. Outcome instruments to measure frailty: a systematic review. Ageing research reviews. 2011. *10*(1):104-114.

12. Saliba D, Elliott M, Rubenstein LZ, Solomon DH, Young RT, Kamberg CJ, Wenger NS. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. Journal of the American Geriatrics Society. 2001. *49*(12):1691-1699.

13. Andrew MK, Mitnitski AB, Rockwood K. Social vulnerability, frailty and mortality in elderly people. *PLoS One*. 2008. *3*(5):e2232.

14. Deleris LA, Yeo GL, Seiver A, Paté-Cornell ME. Engineering risk analysis of a hospital oxygen supply system. Medical decision making. 2006. *26*(2):162-172.

15. Li L, Wang J, Leung H, Jiang C. Assessment of catastrophic risk using Bayesian network constructed from domain knowledge and spatial data. Risk analysis, 2010. *30*(7):1157-1175.

16. Paté-Cornell E, Guikema S. Probabilistic modeling of terrorist threats: A systems analysis approach to setting priorities among countermeasures. Military Operations Research. 2002. *7*(4):5-23.

17. Platis A., Limnios N, Le Du M. (1998). Dependability analysis of systems modeled by non-homogeneous Markov chains. Reliability Engineering & System Safety. 1998. 61(3):235-249.

18. Moayedi BZ, Azgomi MA. A game theoretic framework for evaluation of the impacts of hackers diversity on security measures. Reliability Engineering & System Safety. 2012. *99*:45-54.

19. Guanquan C, Jinhua S. Quantitative assessment of building fire risk to life safety. Risk analysis. 2008. *28*(3):615-625.

20. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web. 1999

21. Gleich DF. PageRank Beyond the Web. SIAM Review. 2015. 57(3):321-363

22. Crisostomi E, Kirkland S, Shorten R. A Google-like model of road network dynamics and its application to regulation and control. International Journal of Control. 2011. *84*(3):633-651.

23. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. 19988. Morgan Kaufmann.

24. Hartfiel DJ. Markov set-chains. Lecture notes in mathematics.1998.

25. Cozman FG. Credal networks. Artificial intelligence. 2000. *120*(2):199-233.

26. World Health Organization. Facts on ageing and the life course. World Health Organization, Geneva, Switzerland. 2*012.* Retrieved from (http://www. who. int/features/factfiles/ageing/en/).

27. ICS 03 080 01. Ability assessment for older adults. 2013. Retrieved from jzshfl.gov.cn/uploads/老年人能力评估.doc.

28. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *The Scientific World Journal. 2001 1:323-336.*

29. Carrière I, Colvez, A, Favier F, Jeandel C, Blain H, EPIDOS study group. Hierarchical components of physical frailty predicted incidence of dependency in a cohort of elderly women. Journal of clinical epidemiology. 2005. 58(11):1180-1187.

30. Spirtes P, Glymour CN, Scheines R. *Causation, prediction, and search. 2000.* Vol. 81. MIT press.

31. Sajja S, Deleris LA. Bayesian Network Structure Learning with Messy Inputs: The Case of Multiple Incomplete Datasets and Expert Opinions. International Conference on Algorithmic Decision Theory 2015. 27:123-138

32. Deleris L, Deparis S, Sacaleanu B, Tounsi L. Risk Information Extraction and Aggregation. International Conference on Algorithmic Decision Theory 2013. 154-166.