Analysis and assessment of injury risk in female gymnastics: Bayesian Network approach

Lyudmila Dimitrova[1], Kristina Petkova[2]

[1]Department of Computer and Information TechnologyUniversity "Prof. As. Zlatarov", Burgas, Bulgaria
[2]Faculty of Industrial technology, Technical University of Sofia , Sofia, Bulgaria

***Abstract* – This paper presents a Bayesian network (BN) model for estimating injury risk in female artistic gymnastics. The model illustrates the connections betweenunderlying injury risk factorsthrough a series ofcausal dependencies. The quantitativepart of the model – the conditional probability tables, are determined using TNormal distribution with parameters, derived by experts. The injury rates calculated by the network are in an agreement with injury statistic data and correctly reports the impact of various risk factors on injury rates. The model is designed to assist coaches and supporting teams in planning the training activity so that injuries are minimized. This study provides important background for further data collection and research necessary to improve the precision of the quantitative predictions of the model.**

***Keywords* – Bayesian network, injury risk analysis, female artistic gymnastics.**

1. Introduction

Gymnastics has always been considered a sport witha high risk of injury [1],[2],[4]. The greater frequency of sport trauma is due to the intensive and physically challenging training of the gymnasts. Gymnastics is a sport that places high physical and psychological demands in terms of strength, flexibility and endurance and is practiced in young age. Published studies on artistic gymnastics report weekly hours of preparation from 7 to 36 hours among females, depending on their competitive level. Top level gymnast typically train twice each day, 5-6 days per week and up to 12 months of the year [2].

Over the years, the difficulty of gymnasts' routines, as well as range of skills required for successful competition, has increased considerably. The difficulty of the sport and the increased amount of practice starting at an early age are likely to be the leading factors for the increased risk of injury. Most competitive gymnasts, and especially those who progress to elite levels, usually incur at least one injury through their sports careers [2].

Depending on the mechanism of injury, they can be chronic (overused) and acute [3]. Acute injuries usually occur suddenly during sports activity as a result of a trauma factor influence. Chronic injuries usually occur after practicing a sport for an extended period of time and are a result from the repetitive action of a trauma factor on one area of the body.

A large number of prospective and retrospective studies in gymnastics injuries, especially for women artistic gymnastics (WAG) have been published[2],[5],[6]. The comparison of injury rates across these studies is challenging because most of them have not taken into consideration the exposure to injury risk, but only reported injury rates per season. Studies reporting injury ratesbased on exposure found rates ranging from 0.5 to 5.3 injuries per 1000 hours of participation[2],[24]. Greece's pre-national team was observed for a year on a weekly basis by Kirialanis at al. in relation to his studies about sport injuries incidence [17]. He recorded acute and overuses syndromes and analyzed the occurrences by sex, age, event and exercise phase. Overall 147 out of 187 athletes experienced acute injuries and 93– overusedinjures. Acute injuries are 61.5 % of all injuries.A closer analysis showed that out of these injuries 29% are mild, 44% - moderate and 29% - major.

Only three studies have reported data relative to athletic exposures (AE) (one AE equals one athlete participating in a competition or practice). Caine at al reported 8.5 injuries per 1000 athletic exposures (AEs) in club level gymnasts [16]. A 16 years study found that the rate of injury in competition was more than twice as high as in practice (15.19 versus 6.07 injuries per 1000 AEs) [10]. Sands et al. reported 90.9 injuries per 1000 AEs in college-level gymnasts [20].

In addition to the personal health risks for the gymnast, the treatment of traumas is associated with substantial expenses. Although it is impossible to eliminate all injuries, efforts for reducing their frequency have positive social as well as economic impacts [21]. An important part of gymnast injury prevention is the identification and analysis of risk factors or factors that contribute to the occurrence of gymnastics injury. Risk factors may be classified as either intrinsic or extrinsic [11],[22],[24]. Intrinsic factors are individual biological and psychosocial characteristics predisposing a gymnast to the outcome of injury. Extrinsic risk factors are factors that have an impact on the gymnast while she is participating in her sport, for example training methods or equipment.

Intrinsic risk factors include physical characteristics, in particular anthropometric characteristics. Factors shown to be associated with increased risk of injury include greater body size, age and body fat, periods of rapid growth [25], [26],[27]. Some data suggest that somatotype may relate to risk of injury. Sleeve at al. reported mesomorphy to be negatively(-) related to injury[28]. Lindner studied the influence of motor characteristics and found that speed (-), balance (-), endurance (+) and flexibility (+) as significant injury predictors among club-level female gymnasts[12],[27].

Relatively unexplored area of injury research in gymnastics is the role of psychological factors in injury occurrence. Kerr at al.reported a moderately strong positive relationship between the number of stressful life events and injury number and severity [29]. Kolt at al. found life stress to be a significant predictor of injury in elite and non-elite competitive female gymnasts [8].

The most important extrinsic risk factor is inappropriate training [2],[3],[23]. According to one study 43.9% of the traumas in artistic gymnastics are due to inadequate training methods, 17.4% - to performance specifics, 9.6% - to training conditions and 8.4% - to the gymnasts condition[19]. Higher rates of injuries among advanced-level gymnasts are observed in previous studies [9],[14],[30]. Relatively high frequency of injuries are establishedto occur during the early part of practice, following periods of reduced training, during competitive routine preparation and during the weeks just prior to competition[15],[28]. Floor exercise is characterized by the highest rate of injury, especially landing [2],[15]. Most of injuries (about 2/3) are acute [7],[13],[18]. National level gymnasts incur relatively high rates of injuries associated with competitions [15],[26].

Despite the great number of studies on injuries in WAG, many unanswered questions and areas of needed research remain [2],[16]. As in every other field involving multiple factors and uncertainty, Bayesian networks can be a helpful tool to integrate different information and in particularly to study injury risk joint distribution utilizing data collected by researchers.

2. Methods and Model

2.1. Bayesian Networks (BN)

A BNis*a directed acyclic graph* (DAG) consisting of nodes representing discrete or continuous random variables linked with directed edges[31],[32],[33]. Thus BN is a graphical representation of uncertain quantities that explicitly reveal the probabilistic causal dependence between the variables and the flow of information in the model. The DAGprovides the qualitative part of causal reasoning in a BN. The quantitative part consist of a conditional probability tables (CPTs) for discrete nodes (or functions for continuous nodes) attached to each node of the network (Fig.1).

The nodes without any edges directed into them are called “root” (“input”) nodes - $A_{1,}$ A_2 in Fig. 1. The CPTs of these nodes are prior probability tables (functions). The nodes that have edges directed into them are called “child” nodes and the nodes from which the edges depart are called ‘‘parent’’ nodes. Each child node has CPT (function),providing the probabilities of its states for all combinations of statesof its parents nodes(Fig.1). If each of *n* parent nodes has *m* states, the given child node has m^{n-1} entries -the size of the CPT grows exponentially as the number of parents increases.

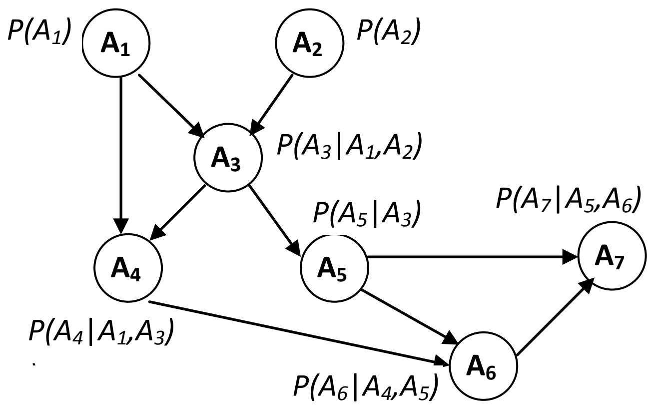

Figure 1.Simple Bayesian Network

According to the chain rule, a Bayesian Network is a presentation of the joint distribution over all the variables represented in the DAG and the marginal and the conditional probabilities can be computed for each node of the network. Let*U* is an universe of variables:

$$U = \{A_1, A_2, ..., A_n\} \qquad (1)$$

The joint probability of distribution P(U) is the product of all conditional probabilities, specified in the BN:

$$P(U) = \prod_{i=1}^{n-1} P(A_i \mid pa_i) \qquad (2)$$

where pa_i is the parent set for A_i. The marginal probability of A_i is:

$$P(X_i) = \sum_{except\, A_i} P(U)) \qquad (3)$$

The BNs main application is answering the probabilistic queries, e.g. determining the posterior distributions of random variables of the BN when some other variables are observed. An evidence is called ‘‘hard’’ when it is an exact observation of the state of the variables, while it is “soft” when a non-definite information is given, expressed in terms of likelihood for the states of the variable [32]. The process of updating the BN given evidence is known probabilistic inference. Updating is consistent with d-separation property of the BN, which characterizes the way in which information flows through different types of node’s connections[34].

Assume the evidence *e* is found, we can calculate probability distribution given the evidence:

$$P(U \mid e) = \frac{P(U,e)}{P(e)} = \frac{P(U,e)}{\sum_U P(U,e)} \qquad (4)$$

One useful aspect of BNs is that there is no requirement for availability of large datasets. The prior conditional probabilities of the model can be estimated by an expert or derived from data using the Expectation-Maximization (EM) algorithm [35], [36], [37]. BNs have a good prediction capability even with small sample sizes [38].

BNs are particularly useful for the analyses of data and expert knowledge in domains that are characterized by uncertainty. There isa large number of applications of BNs in the field of risk analysis [39],[40],[41].

2.2. Description of the problem and proposed model

In our previous works we suggested a BN model for assessing the risk of injury in pre-elite and eliterhythmic gymnasts (RG)[42],[43],[44] . Although many of the underlying risk factors and injury mechanisms are common for both sports, some of them are specific to artistic gymnastics. In addition, while in RG overused injuries are more common, acute injuries are more common in AG[2]. This is due to the difficulty and greater risk associated with AG routines.

The overview of the BN model for estimating injury risk in women AG is presented in a previous study [45]. This paper proposes a BN model for quantitative injury risk assessment per 1000 training hours of female gymnasts competing in high level (elite and sub-elite).

According to the literature gymnasts practice from 7 to 33 hours per week, 20 hours on average [2]. Given that gymnasts rest one month per year, they train for 48 weeks, which totals to 960 hours of training. Therefore the model is correctly assessing the probability for a single gymnast to experience an injury during a year period.

The proposed model is presented in Fig. 2.The network has two end (leaf) nodes: *Overused Injury* and *Acute Injury*- the probabilities of chronic and acute injuries in percentages for 1000h of training, respectively. Due to the substantial differences in the quantitative assessments derived by different studies and since these studies assess only a small amount of the factors included here, the goal of the presented model at this stage is to:

(1) Reflect the cause-and-effect relations between the factors leading to injuries in WAG.
(2) Assess the degree and direction of influence of the factors cited in the studies on overused and acute injuries.
(3) Calculate quantitative estimates by given prior probability distribution of the root nodes to match the cited literature data.

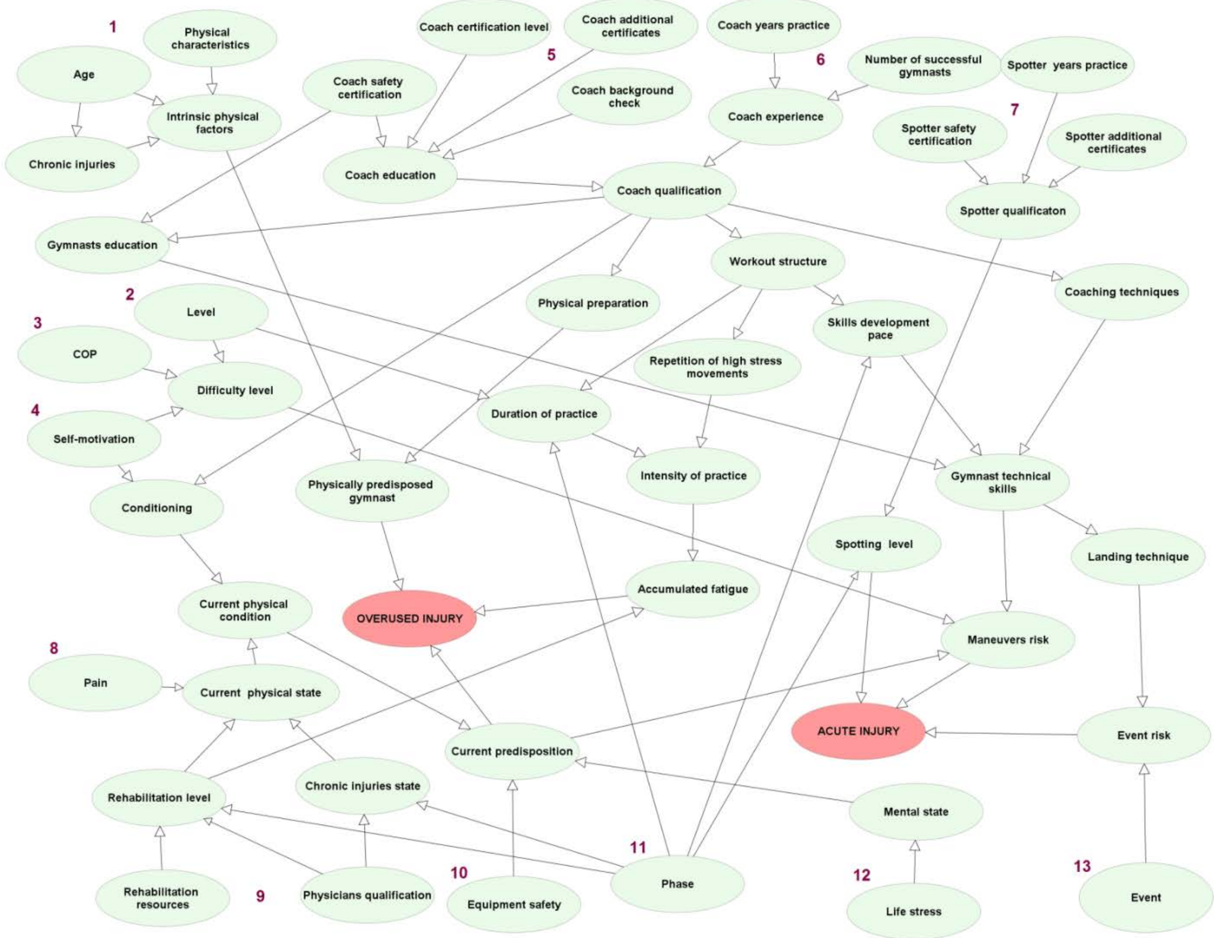

Figure 2.Bayesian Network model for estimating risk of severe overused and acute injuries in WAG

As it is stated above, existing research on similar factors reports results that vary considerably: 0.5 to 5.3 injuries per 1000 hours of training. Thus the goal of this model is to derive quantitative injury risk assessment close to the average of this interval. Such results or overall rate is 2.5 injuries/1000h are reported by Caine andNassar [16]. Kirialanis states about 30% of all injuries are severe, which amounts to around 0.75 severe injuries/1000h[17]. Relative to one gymnast this means an average probability of 75% for experiencing an injury (approximately 50% acute and 25% overused).

To achieve objectives (1)-(3), the structure of the suggested model is synchronized with the mechanisms of injury defined in the literature and the currently available statistical data for factors influencing the injury rate that has been usedtoevaluatethe model parameters.

In the current version, three possible states exist for each node.Design and reasoning with the model is done with AgenaRisksoftware[46].

The following parameter groups are included as underlying factors (i.e. root nodes) of the network (Fig. 2):

1. Parameters defining intrinsic characteristic of the athlete, her aptitude for gymnastics activities: *Age*, *Physical characteristics* and presence*ofChronicinjuries*.Physical characteristics are determined by parameters such as height, weight, flexibility, etc., whose influence is currently not sufficiently studied [18].
2. The level of athlete - (node *Level*), which largely determines the intensity of the workload and difficulty of her program.
3. *Code of Points*, which becomes more complicated with every Olympic cycle.
4. Motivation of the gymnast for high sport achievements and medals – node *Self-motivation.*
5. Parameters defining the specialized theoretical knowledge of the coach (node *Coach Education*). The indicators for assessing this network variable can differ by country. The presented model includes some of the possible indicators in the USA: *CoachCertification Level, Coach Safety Certification, Additional Certificates, and Background Check.*
6. Parameters defining the positive experiences of the coach (*Coach Experience*) in training high

performing gymnasts: *Years Practice* and*Numberof successful gymnasts* per year of coaching experience.The level of education andexperience defines the qualification of the coach (node *Coach Qualification*), which plays a key role in gymnasts preparation as well as in minimizing the injury risks.

Since artistic gymnastics involves difficult coordination, various techniques are important for its safe practice. Of particular importance is spotting:a safety technique, designed to assist with the execution of the skill and reduce the possibility ofinjury [2], [23]. The effectiveness of spotting is presented by the node *Spotter Qualification*, which in our model depends on values of the next three indicators:

7. Spotters *Safety Certification, Spotters Additional Certificates and Spotters Years Practice.*
8. Node *Pain*that determines the current physical condition of the gymnast.
9. Indicators for the quality of rehabilitation and medical support – nodes *Rehabilitation Resources* and *Physicians Qualification.*
10. Use of safety devices and equipment - node *Equipment safety.*
11. The preparation phase - node*Phase* .
12. Psychological factors - node*Life Stress.*
13. Competition components – node *Event* (vault, uneven bars, balance beam, floor exercise)

The influence of the listed factors on the outputs of interest – *Overused Injury*and *Acute Injury*occurs via active trails, defined by the model. For example, on coachqualification depend: (a) the quality of the physical preparation of the gymnast (node *Physical Preparation*), (b) the training structure *(Workout Structure)* with components *Duration*, *Repetition of High Stress Movements and Skills Development Pace*, (c) theoretical preparation of the gymnast for performing the routine (*Gymnasts Education*) and (d) the coach ability to provide the gymnast with adequate technical training (*Coaching Techniques*). Therefore, the coach's qualification has an influence on both overused injuries and acute injuries (Fig.2).

According to the literatureoverused injuries are a result from fatigue and substandard physical condition and do not depend on the difficulty of the performed elements [3]. In the presented model, three variables influence the risk for overused injury (Fig. 3): the physical predisposition of thegymnast (*Physically Predisposed Gymnast*), fatigue accumulated over time (*Accumulated Fatigue*) and the individual predisposition of the gymnast at the given moment (*Current Predisposition).*

The first of these factors - *Physically Predisposed Gymnast* characterizes the level of preparation that has been achieved throughout her sports career and is dependent on her abilities (*Intrinsic Physical Factors*) as well as preparation (*Physical Preparation),* contributed by the coach. *Accumulated fatigue* is defined by the overloading during the current training seasonand depends on the intensity of practice as well as on the rehabilitation level. *Current Predisposition* defines the risk of injury for the specific workout and depends on *Current Physical Condition*, as well as *Mental State* as well as the safety measures taken (*Equipment Safety*) (Fig. 3).

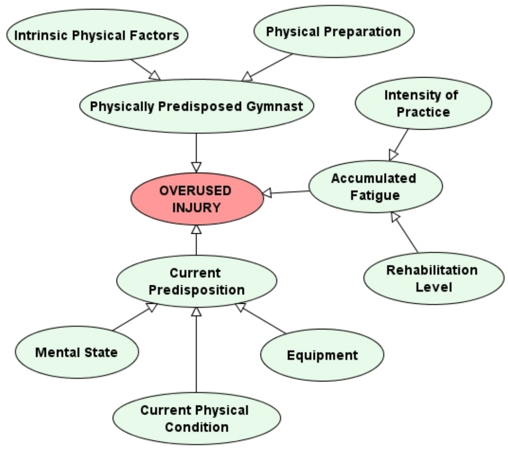

Figure 3. Part of the model, estimating overused injuries rate in WAG

As discussed earlier, the higher frequency of gymnasticsacute injuries is due to the extremely difficult and potentially dangerous elements performed by the gymnasts. Particularly high is the risk associated with floor exercise landing [2]. In order to incorporate these facts from the literature in the proposed model the node ACUTE INJURYhas three direct parents: *Maneuvers Risk*, *Event Risk and Spotting Level*(Fig.4)**:**

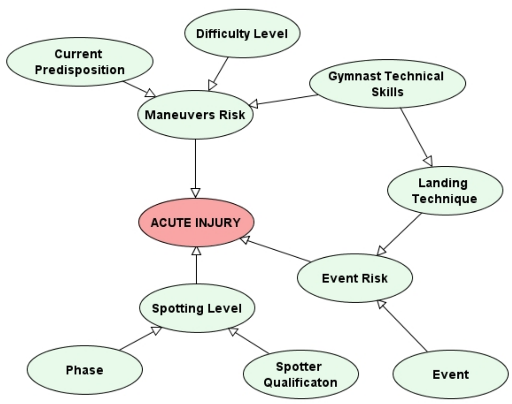

Figure 4 Part of the model, estimating acute injuries rate in WAG

Maneuvers risk defines the risk of performing the specific element and depends on *Current Predisposition* to injury (acute or overused), *Difficulty Level*of thegymnasts programand *Gymnast Technical Skills*. *Event Risk* defines the influence of a specific exercise (node *Event*) and *Landing Technique*. *Spotting Level* depends on spotter qualification and phase - during the competitive period the aid from spotters and improved safety measures are less substantial.

Prior probability distributions of the root nodes are approximately evaluated. CPT for intermediate and leaf nodes are calculatedwith the use of functions.Therefore in the proposed model all of the nodes are defined asRanked in AgenaRisk. Ranked nodes are mapped to underlying numerical scale. This means that no matter what the state labels are and how many states a node has, there is an assumption that there is an underlying numerical scale that goes from 0 to 1 in equal intervals. By defining theranked scale the user must be sure that the labelling of the states is consistent from worst to best. In our case in the beginning of the scale should be the states of the nodes that lead to the higher risk of injury. For example, to model *Accumulated fatigue* in terms of factors *Intensity of practice* and *Rehabilitation level* in consistent way, states of the *Intensity of practice* should be ordered from highest to lowest value, and states of *Rehabilitation level* – from worst to best level.

The underlying numerical scale of the ranked nodes allows defining numerical statistical distribution expressions. Especially useful for defining NPT is the truncated Normal distribution (TNormal). Unlike the regular Normal distribution TNormal has finite end-points. For ranked nodes these endpoints are 0 and 1, respectively. TNormal(like Normal) is characterized bytwo parameters: mean and variance, and allows modeling differentdistribution shapes. Special cases ofTNormal are the uniform distributions, achieved when variance is very high, and the highly skewed distributions, achieved when the mean is close to one of the ends of the range and variance is close to 0.

Due of its properties Tnormalis flexible and capable to generate satisfactory CPT for ranked nodes with ranked parents. The mean parameter can be set to take into account the severity of the impact of a given parent on the probability distribution of the child node. We accomplish thisby using weighted expressions for the mean parameter. The parent nodes with higher weight have a greater influence. In the simplest case the parameter mean is determined as a weighted mean of the parent nodes:

$$WMEAN = \frac{\sum_{i=1\ldots n} w_i X_i}{n} \qquad (5)$$

Here $w_i \geq 0$ are weights, and n is number of parent nodes. In AgenaRisk syntax of the function is:

$wmean(w_1,\ parent_1,\ w_2,\ parent_2, \ldots,\ w_N, parent_N)$ (6)

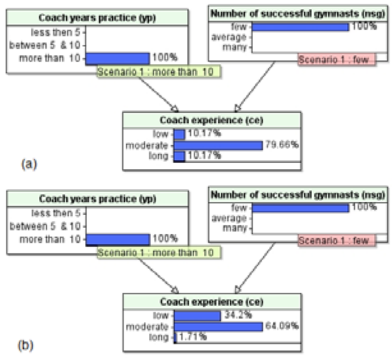

Figure 5 Eliciting CPT using wmean function with different weights of parent nodes
(a) *ce=TNormal(wmean(1,yp, 1, nsg), 0.01)*
(b) *ce=TNormal(wmean (1, yp, 2, nsg), 0.01)*

Fig. 5 demonstrates the use of *wmean* in creating the CPT of the node *Coach experience* with parent nodes *Coach years practice (yp)* and *Number of successful gymnasts (nsg).* If we assume that the two factors (*yp* and *nsg*) have the same weight in the assessment of *coach experience*, the resulting distribution is presented onFig.5, a. In order to assign a greater importance to coach experience *number of successful gymnasts,* we should use a*wmean*with a greater weight of the*nsg*(Fig.5 ,b).

The shift of the distribution to one of the ends can be achieved using the *wmean*with correctioncoefficient. In the presented example we use coefficient <1 (0.8).This setting allows deriving lower probabilities for the high qualification of the coach in preventing injuries despite their long work experience and coaching a high number of successful gymnasts.

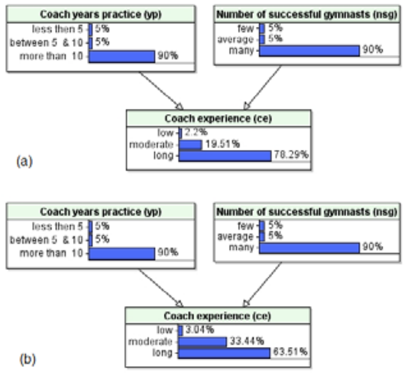

Figure 5 Eliciting CPT using wmean function with correction coefficient
(a) *ce=TNormal(wmean(1,yp, 2, nsg), 0.01)*
(b) *ce=Tnormal(0.8*wmean (1, yp, 2, nsg), 0.01)*

If a simple weighted mean (WMEAN) does not satisfy the requirements for the distribution, the built in AgenaRisk functions WMIN, WMAX and

MIXMINMAX can be used. The *weighted min* function is defined as:

$$WMIN = \min_{i=1...n}\left[\frac{w_i X_i + \sum_{i\neq j} X_j}{w_i + (n-1)}\right] \qquad (7)$$

The analogous *weighted max* function has the following general form:

$$WMAX = \max_{i=1...n}\left[\frac{w_i X_i + \sum_{i\neq j} X_j}{w_i + (n-1)}\right] \qquad (8)$$

Finally, the function MIXMINMAX is a mixture of two functions WMIN and WMAX with weights w_{min}andw_{max}, respectively

$$MIXMINMAX = \frac{w_{min}\, WMIN + w_{max}\, WMAX}{w_{min} + w_{max}} \qquad (9)$$

In AgenaRisk syntax of the function is

mixminmax(w_{min},w_{max},parent$_1$,parent$_2$,..,.parent$_N$)
(10)

The following Figure 6 illustrates the use of these three functions. WMIN shifts the distribution towards the beginning of the scale and WMAX towards the end. MINMINMAX combines the two functions with the predefined weights.

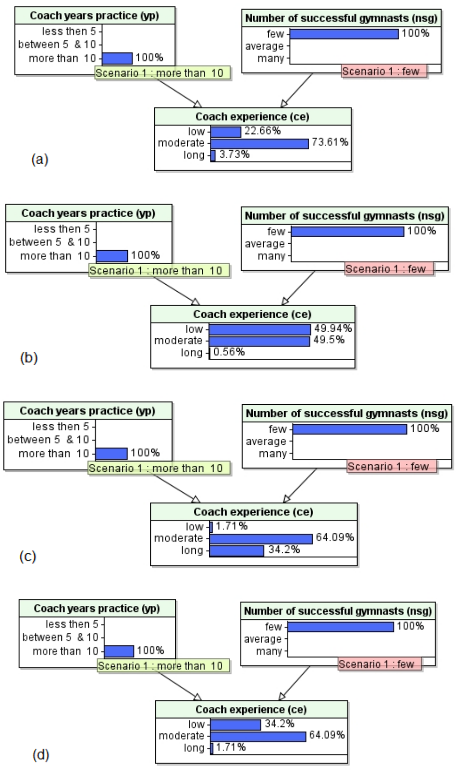

Figure 6 Eliciting CPT with different functions

(a) ce=TNormal(wmean(1,yp, 2, nsg), 0.01)
(b) ce=Tnormal(wmin(1,yp, 2, nsg), 0.01)
(c) ce=Tnormal(wmax(1,yp, 2, nsg), 0.01)
(d) ce=Tnormal(mixminmax(1,2,yp, nsg), 0.01)

In AgenaRisk we can use partitioned expressions as well in order to calculate CPT with different functions for different states of the given node. To do this we must select what parent node to condition on (Fig. 7). In the provided example this allows deriving different probability distributions for different age categories.The resulted distributions are shown in Figure 8.

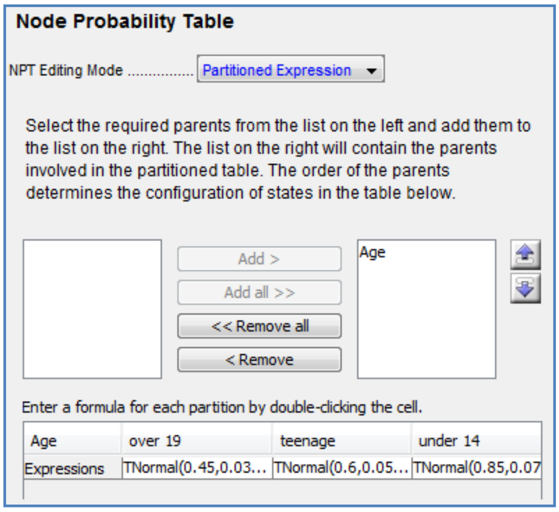

Figure 7. Selecting the CPT(Node Probability Table in AgenaRisk) for the node Chronic Injury as partitioned expression conditional on „Age"

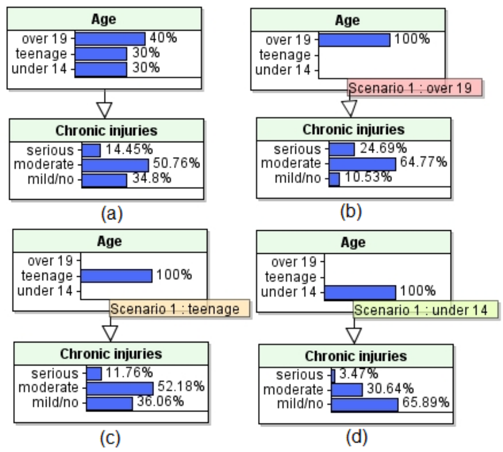

Figure 8.Calculated CPT for node Chronic Injury with partitioned expression conditional on „Age"

In the proposed model the desired distributionsare achieved throughthe usage ofappropriate functions to evaluate the mean parameter ofTNormaland by setting the weight of different factors involved in the selected function. Uncertainty in the evaluation of CPT is set via the variance parameter of TNormal. Model parameterization is presented in the tables 1-8.

*Table 1.Properties of the node**intrinsicphysical factors** and its parents*

Node name	Node ID	Node states	Node CPT
Age	*age*	*over 19* *teenage* *under 14*	(0.40, 0.30, 0.30)
Physical characte-ristic	*pc*	*average* *very good* *excellent*	(0.20, 0.40, 0.40)
Chronic injuries	ci	*serious* *moderate* *mild/no*	age=”*over 19*”: TNormal(0.45, 0.03) age=”*teenage*”: TNormal(0.6, 0.05) age=”*under 14*”: TNormal(0.85, 0.07)
Intrinsic physical factors	*ipf*	*low* *medium* *high*	*TNormal* *(wmean(1.0, age, 1.5, pc, 1.5, ci), 0.01)*

*Table 2.Properties of the nodes **spotting level,maneuvers risk** and **event risk** andtheir direct and indirect parents*

Node name	ID	Node states	Node CPT
Spotting level	*sl*	*low* *medium* *high*	*TNormal(* *wmean(1.0, sq, 1.0, ph),* *0.0005)*
Skill development pace	*sdp*	*very fast* *fast* *optimal*	*TNormal(* *wmean(2.0, ws, 1.0, ph),* *0.0005)*
Gymnasts technical skills	*gts*	*good* *very good* *perfect*	*TNormal(wmean(3.0, ct,* *1.0, ge, 1.0, sdp),* *0.0005)*
Landing technique	*lt*	*good* *very good* *perfect*	*TNormal(wmean(1.0, gts),* *0.0005)*
Maneuvers risk	*mr*	*very high* *high* *moderate*	*TNormal(wmean(3.0, dl,* *1.0, gts, 1.0, cp),* *0.0005)*
Event	*ev*	*floor* *fault* *uneven bars* *beam*	*(0.25, 0.25, 0.25, 0.25)*
Event risk	*er*	*very high* *high* *moderate*	*event=”floor”:* *TNormal(1.3* lt,0.0005)* *event=”vault” :* *TNormal(1.2* , lt, 0.0005)* *event=”uneven bars”,” b. beam”* *TNormal(lt, 0.0005)*

*Table 3. Properties of the nodes **physically predisposed gymnast** and **accumulated fatigue** and their direct and indirect parents (not listed before)*

Node name	ID	Node states	Node CPT
Physical preparation	*pp*	*good* *very good* *excellent*	*TNormal(cq, 0.005)*
Physically predisposed gymnast	*ppg*	*high* *moderate* *low*	*TNormal(1.0, ipf, 1.0, pp),* *0.0005)*
Workout structure	*ws*	*non-optimal* *good* *excellent*	*TNormal(cq, 0.0005)*
Duration of practice	*du*	*very long* *extended* *optimal*	*TNormal(wmean(2.0, ws,* *1.0, ph, 2.0, le),* *0.0005)*
Repetition of high stress movements	*rhm*	*high* *medium* *low*	*TNormal (ws, 0.0005)*
Intensity of practice	*ip*	*very high* *high* *optimal*	*TNormal(wmean(1.0, du,* *1.0, rhm) , 0.0005)*
Accumulated fatigue	*af*	*high* *moderate* *low*	*TNormal(wmean(1.0, ip,* *1.0, rl), 0.0005)*

*Table 4. Properties of the nodes **coachqualification,coaching techniques** and **gymnasts education** and their direct and indirect parents*

Node name	ID	Node states	Node CPT
Coach certification level	*ccl*	*level 1* *level 2* *level 3*	*(0.20, 0.50, 0.30)*
Coach safety certification	*csc*	*no* *good* *excellent*	*(0.20, 0.30, 0.50)*
Coach additional sertificates	*cas*	*no* *good* *excellent*	*(0.20, 0.30, 0.50)*
Coach background check	*cbc*	*no* *good* *excellent*	*(0.30, 0.30, 0.40)*
Coach education	*ced*	*low* *medium* *high*	*TNormal (wmean(4.0, ccl,* *1, csc, 2.0, cas, 2.0, cbc),* *0.005)*
Coach years practice	*cyp*	*< 5* *5-10* *>10*	*(0,20, 0.20, 0.60)*
Number sussesful gymnasts	*nsg*	*few* *average* *many*	*(0.20, 0.50, 0.30)*
Coach experience	*cex*	*low* *moderate* *long*	*TNormal (* *wmean(2.0, nsg, 1.0, cyp),* *0.001)*
Coach qualification	*cq*	*basic* *high* *very high*	*TNormal (* *wmean(1.0, ced, 1.0, cex),* *0.0005)*
Coaching techniques	*ct*	*good* *very good* *excellent*	*TNormal (cq, 0.0005)*
Gymnasts education	*ge*	*low* *medium* *high*	*TNormal (* *wmean(1.0, csc, 2.0, cq),* *0.01*

*Table 5.Properties of the node **difficulty level** and its parents*

Node name	ID	Node states	Node CPT
Level	*le*	*top* *very high* *high*	*(0.40, 0.30, 0.30)*
Code of points	*cop*	*very traumatic* *traumatic* *slightlytraum.*	*(0.80. 0.15, 0.05)*
Self motivation	*sm*	*very high* *high* *average*	*(0.20, 0.40, 0.40)*
Difficulty level	*dl*	*very high* *high* *average*	*TNormal (wmean(3.0, le,* *2.0, sm, 1.0, cop),* *0.001)*

*Table 6. Properties of the nodes **current predisposition** and its direct and indirect parents (not listed before)*

Node name	*ID*	*Node states*	*Node CPT*
Conditioning	*co*	*unsufficient* *good* *excellent*	*TNormal(wmean(1,cq,* *1.5, 1-sm),* *0.005)*
Pain	*pa*	*intense* *bearable* *mild/no*	*(0.10, 0.50, 0.40)*
Phase	*ph*	*competitive* *pre-competitive* *preparatory*	*(0.50, 0.30, 0.20)*
Rehabilitation resources	*rr*	*good* *very good* *excellent*	*(0.10, 0.20, 0.70)*
Physictians qualification	*pq*	*good* *very good* *excellent*	*(0.10, 0.30, 0.60)*

Rehabilitation level	*rl*	*incomplete* *good* *excellent*	*TNormal(wmean(2.0, rr, 0.5, ph, 1.5, pq), 0.0005)*
Chronic injuries state	*cis*	*activated* *partially cured* *cured*	*TNormal(mixminmax(1, 3, pq, ph), 0.0005)*
Current physical state	*cps*	*painkillers use* *good* *excellent*	*Pain="mild/no"* *TNormal(wmean(1.0, rl, 1.0, cis), 0.0005)* *Pain="intense", "bearable"* *TNormal(wmean(1.0, rl, 1.0, cis, 2.0, pa), 0.0005)*
Current physical condition	*cpc*	*non-optimal* *good* *excellent*	*TNormal(wmean(1.0, co, 2.0, cps), 0.0005)*
Equinpment safety	*es*	*conditions on competiton* *partially sequred* *fully secured*	*(0.10, 0.30, 0.60)*
Life stress	*ls*	*high* *medium* *low*	*(0.10, 0.30, 0.60)*
Mental state	*ms*	*anxious* *good* *excellent*	*TNormal(1.5*wmean(1.0, ls), 0.0005)*
Current predisposition	*cp*	*high* *moderate* *low*	*TNormal(wmean(3.0, cps, 1.0, es, 1.0, ms), 0.0005)*

*Table 7.Properties of the node **spotter qualification**and its parents*

Node name	ID	Node states	Node CPT
Spotter safety certification	*ssc*	*no* *good* *excellent*	*(0.05, 0.25, 0.70)*
Spotter additional sertificates	*sas*	*no* *good* *excellent*	*(0.20, 0.30, 0.50)*
Spotter years practice	*syp*	*< 5* *5-10* *>10*	*(0,30, 0.30, 0.40)*
Spotter qualification	sq	*basic* *high* *very high*	*TNormal (wmean(2.0, ssc, 1.0, sas, 1.0, syp), 0.0005)*

CPTfor node*Overused injury* is calculated with *TNormal(wmean(1.0, ppg, 1.0, af, 1.0, cp), 0.0005),* i.e. using *wmean*function with equal weights of parent nodes *Physically Predisposed Gymnast* (*ppg*), *Accumulated Fatigue (af)* and *Current Predisposition(cp).*

Similarly, acute injury CPT (with parent nodes *Spotting Level (sl) , Maneuvers Risk(mr), Event Risk(er) is* computed as*TNormal(wmean(1.0, sl, 1.0, mr, 1.0, er), 0.0005).*

2.3. Results and discussion

Figure 9 demonstrates marginalprior probability distribution in the model for the estimated prior probabilities of the root nodes. The model calculates 23.82% probability of the state „*yes*" for the node *Overused injury* and 38.38% probability of the state „*yes*" for the node *Acute injury*. The ratio*Acute injurues="yes"/Overused injury="yes"*is 1.61. This assessment is consistent with the data by Kirialanisat al. cited previously, which concluded a ratio of *Acute injuries / Overused injuries* = 61.5%/38.5% = 1.60[17].

As previously clarified, when calculated for an individual gymnast, this prediction translates to 62.2% probability for trauma in 1000 hours of training or in 1 year by training 20 hrs/week or even sooner in case training is more intensive.

If applied to a group of gymnasts, the model predicts that after 1000 hours of training 2 out of 10 gymnasts will suffer from a serious overused injury (time loss>30 days). Respectively, 4 out of 10 gymnasts will suffer fromsevereacute injuries.

The advantage of the model is that it allows an assessment of this probability depending on the values of *all contributing factors.* The mechanism of their influence can be further understood after acquiring additional observational data, as well as structure learning of the model. Additional data is necessary especially as it concerns coach qualification. It is well known that trainers have substantial influence on the quality of the training process. The available data doesn't allow an analysis of this important indicator.

Figures 10 and 11 demonstrate the results from sensitivity analysis withAgenaRisk[46] asTornado graphs. *Acute Injury*and*Overused Injury* are defined astarget nodes of the sensitivity analysis. The two charts demonstrate the influence of 14 sensitivity nodes on target nodes *Acute injury="yes"* and *Overused injury="yes"*, respectively. The length of the rectangular for sensitivity node determines the degree of influence of this node on the respective target node.

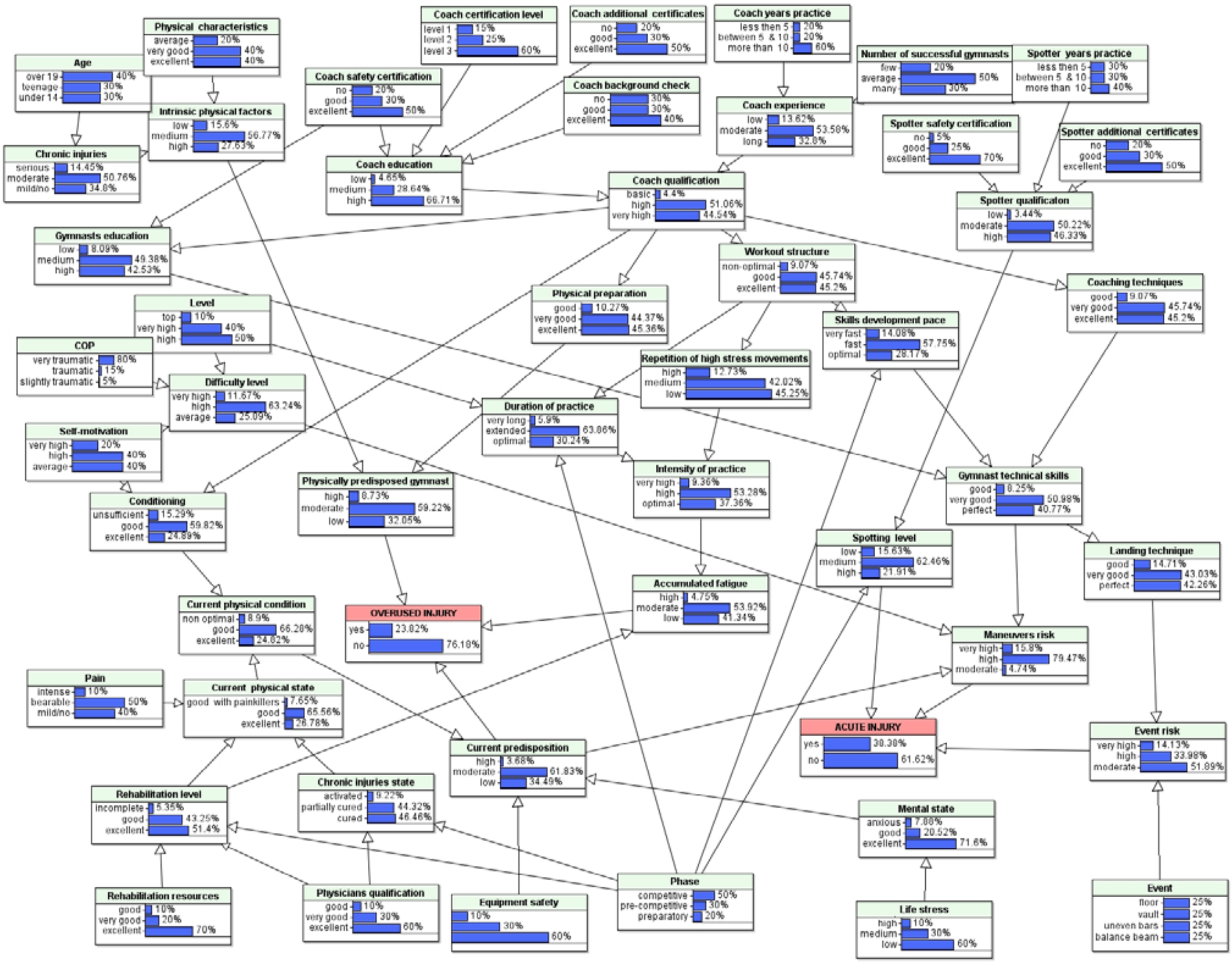

Figure 9. Prior marginal probability distribution in the proposed model

As demonstrated on Fig.10, the factors with most influence on*Acute injury=”yes*”are: (1) Spotter qualification; (2) Coach education; (3) Coach experience; (4) Gymnasts level; (5) Event; (6) Code of points (COP); (7) Equipment safety.

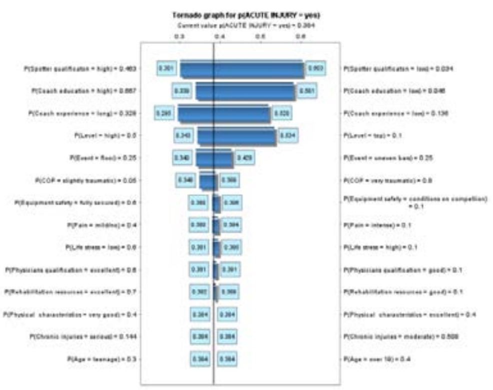

Figure 10. Impact of the sensitivity nodes on the target node Acute injury value “yes”

This is in agreement with the logic underlying the model: acute injuries arespecific for artistic gymnastics and result from the great difficulty and risk of the performed elements. The role of spotters is well known and, as we know, the gymnasts use it often for especially difficult and not well rehearsed elements. The role of the trainer in reducing acute injuries is to assure an adequate adjustment of the speed in learning new elements and correct technical preparation (including landing technique) of the gymnasts. Gymnastslevel и COP determine the difficulty and injury risk of the elements. In some situations the risk of acute trauma can be reduced by using safety equipment. The model doesn’t reflect a relationship between acute injuries and the age of the gymnast. The eventualcausal relationship between the two nodes can be defined when statistical data become available.

In contrast to acute injuries, overused injuries due to accumulated fatigue as a result from ongoing and highly intensive training.Overused injuriesare mostly influenced by the following factors(Fig. 11):(1) Coach education; (2) Coach experience; (3) Rehabilitation resources; (4) Physicians qualification; (5) Chronic injuries; (6) Age; (7) Equipment safety; (8) Physical characteristics;

(9) Life stress; (10) Pain; (11) Level of the gymnasts.

Trainers play a role in overused injuriesby appropriate planning of the workout structure and the individual assessment of the duration of trainings – as determined by the model structure.The influence of modern rehabilitation resourcesand highly qualified specialists in maintaining a good gymnast condition is without a doubt critical. Overused injuries are affected bychronic injuriesand pain as an indicator of their development, as well as by the physical characteristics and age of the gymnast – with age some characteristics of the tissues such as elasticity decline. The use of safety equipment leads to reducing the level of stress as well as fatigue.

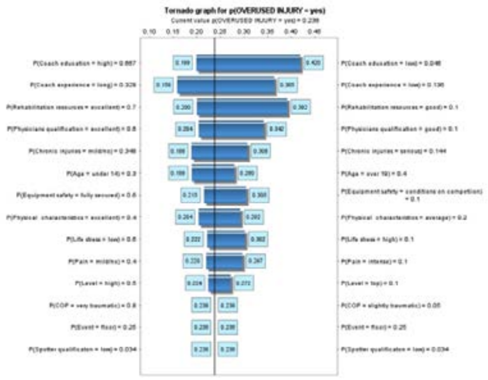

Figure 11. Impact of the sensitivity nodes on the target node Overused injury value "yes"

Evidential reasoning with the proposed model can be used by the training teams and gymnasts as a tool for buildingtraining and competition plans in order to minimize injuries. By defining the combination of node values present at a given moment, the posterior probability of injury rates can be calculated.

For example, Figure12shows posterior marginal probability distribution for who possible observations of the some input and intermediate nodes - scenarios 2 and 3 in table 8(Scenario 1 is prior probability distribution).

Table 8.Observations for two possible scenarios

Node	***Scenario 2***	***Scenario 3***
Age	Over 19	teenage
Coach certification level	Level 2	Level 3
Coach years practice	5-10	Over 10
Chronic injuries	serious	mild/no
Pain	intense	nild/no
Level	top	top
Phase	competitive	pre-competitive
Rehabilitation. resources	good	excellent
Equipment safety	conditions on competition	fully secured

For unfavourable scenario 2 model gives 72.23% probability of *Acute injury="yes"* and76.37% probability of *Overused injury="yes"*. In a more favorable scenario 3, the model calculates 40.47% probability of *Acute injury="yes"* and10% probability of *Overused injury="yes"*.The substantial change in the trauma risk is consistent with the greater amount of statistical data. The presented model allows an analysis of the steps leading to reducing the injury risk. Intop level gymnasts, as in the selected scenarios, the exercises are more difficult and there are, consequently, high risk of acute injuries in both cases. There are,however greater possibilities for reducing overused injuries. The model demonstrates that the participation in training and competitions in the presence of chronic trauma and pain leads to a 7-fold increase of the risk for overused injury compared to a scenario of healed trauma and lack of pain.

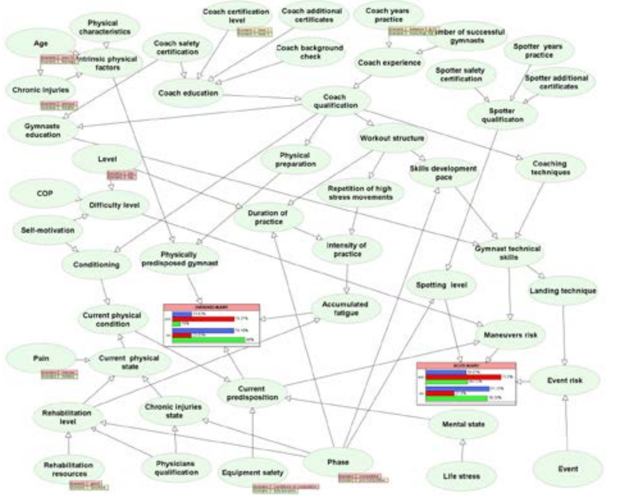

Figure 12.Prior marginals (blue) and posterior marginals of the end nodes for scenarios 2(red) & 3(green)

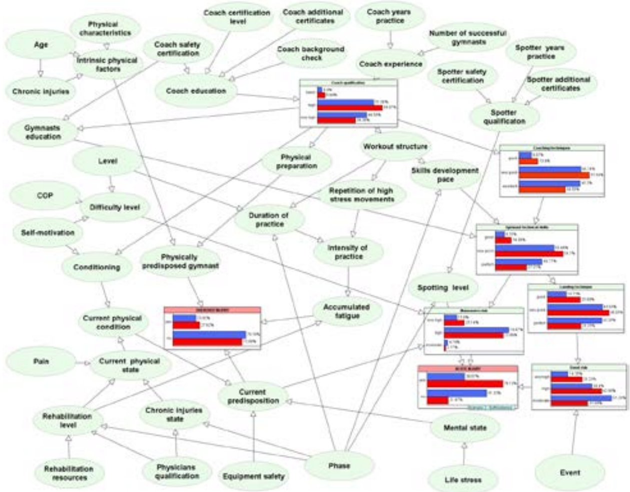

Figure 13. Some prior (blue) and posterior (red)marginals for given soft evidence (0.85,015) of the node Acute injury

The model can also be used for causal reasoning (analysis of the reasons for increased injury rate). For example, Fig.13 shows marginal posterior probability distribution in the presence ofsoft evidence (0.85, 0.15) for*Acute Injury* node. The change of posteriormarginals of some input and intermediate nodes can be used to analyze the causes leading to this high injury rate. For the given soft evidence *Coach Qualification="very high"* changes fromprior value 44.54% to 35.02% - i.e.one of the reasons for the observed higher injury rate could be underqualification of the coaches. As is shown on Fig. 13, significant changes are observed in the marginal probability distribution of the nodes*Coaching techniques, Gymnasts Technical skills, Landing Techniques, Maneuvers Risk, Event Risk.*

3. Conclusion

Artistic gymnastics is a popular and entertaining sport. Achieving a high level of proficiency in artistic gymnastics requires intensive and specialized training that often results in serious injuries.

A number of previous studies have analyzed the risk factors for traumas, the mechanisms of their occurrence as well as pathways to prevent them. The summarizing, analyzing and utilizing such volume of information is very challenging without the use of artificial intelligence. This is a first attempt to link the quantitative data on injury in artistic gymnastics with the mechanisms of injury occurrence. The proposed a Bayesian Network model can be used for assessing the influence of multiple factors on the injury risk as well as for analysis of the factors leading to increased injury levels.

The model allows a quantitative assessment of the injury risk depending on the expert-defined parameters about the training process. In this sense, the model can be a useful too, particularly among less experienced coaches.

Among our major goals was to draw attention to the capabilities of BN as a tool for assessment and diagnosis of injury risk. As with all artificial intelligence models, the accuracy of the assessments can be improved through achieving a greater precision of the model parameters and improving the structure of the model using statistical data. Dueto the great variability in the available statistical data, as well as the inconclusive and contradictive evidence regarding some risk factors, the quantitative estimates for injury rate cannot be entirely precise at this stage.

One important issue highlighted by this work is the lack of standardization in the methodology for documenting and analyzing injury risk an artistic gymnastics. In accordance with the work of Caine and colleagues [2], this work demonstrates that there is a pressing need for a FIG consensus statement that identifies injury definitions and research methodology.

The presented BN model provides a solution to the problem of the lack of systematic collection of statistical data by various research groups. In order for the statistical data to be compatible, injury data records should contain values for as many as possible of the parameters incorporated in the model. More consistent data and the use of such data for parameter and structure learning of the model will increase its value as a tool for causal and evidential reasoning.

References

[1]. Kolt, G., & Caine, D. (2007). *Gymnastics, Epidemiology of Injury in Olympic Sports*, Wiley-Blackwell

[2]. Caine, D., Russell, K., & Lim, L. (2013). *Handbook of Sports Medicine and Science, Gymnastics*, Wiley-Blackwell

[3]. Dubrpvskij, V., (2002). *Sportivnaya medicina*, Vlados

[4]. Snook, G., (1979), "Injuries in women's gymnastics. 5 year study", *Am J Sports Med*. , 7(4), 242-244

[5]. McAuley,E., Hudash,G., Shields,K., Albright,J., Garrick, J., Requa,R., & R.K. Wallace, R. (1988) "Injuries in women's gymnastics. The state of the art", *Am J Sports Med*, 16, Suppl.1, 124-131

[6]. Garrick,J.,&Requa,R., (1980). "Epidemiology of women's gymnastics injuries, *Am J Sports Med*. 8(4), 261-264

[7]. Kolt, G., &Kirkby, R. (1995). "Epidemiology of injury in Australian female gymnasts", *Sports Medicine., Training and Rehabilitation*., 6(3), 223-231

[8]. Kolt, G., &Kirkby, R.,(1996)."Injury in Australian female competitive gymnasts: A psychological perspective",*Aust J Physiotherapy*., 42(2), 121-126

[9]. Kolt, G., &Kirkby, R.,(1999)."Epidemiology of injury in elite and subelite female gymnasts: a comparison of retrospective and prospective findings", *Br.J Sports Med*, 33 (3), 312-318

[10]. Lowry, C., & Leveau,B., (1982). "A retrospective study of gymnastics injuries to competitors and noncompetitors in private clubs", *Am J Sports Med*,10 (4), 237-239,

[11]. Wadley,G., &Albright,J., (1993). "Women's intercollegiate gymnastics. Injury patterns and "permanent" medical disability", *Am J Sports Med*,. 21(2), 314-320

[12]. Lindner,K., &Caine,D., (1990). "Injury patterns of female competitive club gymnasts", *Can.J Sports Sci*., 15(4), 254-261

[13]. Bak,K., Kalms,S., Olesen,S., Jargensen, U., (1994). "Epidemiology of injuries in gymnastics", *Scand J Med & Sci in Sports*, 4(2), 148-154

[14]. Caine,D., Cochrane, B., Caine, C., & Zemper, E., (1989). "An epidemiologic investigation of injuries affecting young competitive female gymnasts", *Am J Sports Med*, 17 (6), 811-820

[15]. Caine,D., Knutzen,K., Howe,W., Keeler,L., Sheppard,L., et al. (2003). "A three-year epidemiological study of injuries affecting young female gymnasts ", *Phys Ther in Sport*,4(1), 10-23

[16]. Caine,D., &Nassar,L., (2005). "Gymnastics injuries", *Med Sport Sci*, 48, 18-58

[17]. Kirialanis,P., Malliou,P., Beneka,A., Gourgoulis,V., Giofstidou,A., & Godolias,G., (2002). "Injuries in

artistic gymnastic elite adolescent male and female athletes, *J Back Musculoskelet Rehabil*,.16(4), 145-151

[18]. Caine,D., Caine,C., &Maffuli, N., (2006). "Incidence and distribution of pediatric sport-related injuries", *Clin J Sport Med*, 16(6), 501-514

[19]. Zaitzev, V. „Travm mozno izbezat", *Gimnastika*, vol.2, pp.44-45, 1986

[20]. Sands,W., Shultz,B., &Newman,A., (1993) "Women's gymnastics injuries: 5 - year study, *Am J Sports Med*, 21(2), 271-276

[21]. Kisser, R., & Bauer R., (2012). "The Burden of Sport Injuries in the European Union",[Online] http://www.eurosafe.eu.com/csi/eurosafe2006.nsf

[22]. Vanderlei,F., Marques Vanderlei, L., Júnior,J., & Pastre, C., (2013). "Characteristics of sports injuries and factors associated with injury in beginners of female artistic gymnastics", Fisioter. Pesqui., 20 (2), Online: http://www.scielo.br/scielo.php

[23]. Daly,R., Bass,S., &Finch,C., (2001) "Balancing the risk of injury to gymnasts: how effective are the counter measures?", *Br J Sports Med*, 35, 8–20

[24]. Freddie,H., David,A., & Stone,M., (1994). "Sports injuries: Mechanisms, prevention, treatment"*Gymnastics, 2nd edition, Chapter 26,* Lippincott, Williams & Wilkins

[25]. Schneider,S., Yamamoto,S., Weidmann,C., &Brühmann,B., (2012). "Sports injuries among adolescents: incidence, causes and consequences", J Paediatrics & Child Health, 48(10), 183-189

[26]. Colvin,A., &Lynn,A., (2010). "Sports-related injuries in the young female athlete", *Mt Sinai J Med.*, 77(3), 307-314

[27]. Lindner,K., &Caine,D., (1993). "Physical and performance characteristics of injured and injury-free female gymnasts", *J Hum Mov Stud*, 42, 121-126

[28]. Steele, V., &White,J., (1986). "Injury prediction in female gymnasts". *Br J Sports Med* , 20, 31-33

[29]. Kerr,C., & Minden, H., (1988). "Psychological factors related to the occurrence of athletic injuries", *J.Sport. Exer.Psych.* 10(2), 167-173

[30]. Mackie,S., &Taunton,J., (1994), "Injuries in female gymnasts.Trends suggest prevention tactis", *Phys. Sportmed*, 22, 40-45

[31]. Heckerman,D., (1997). "Bayesian Networks for Data Mining", *Data Mining and Knowledge Discovery*, 1(1), 79-119

[32]. Kjaerulff,U.,&Madsen,A., (2005). "An Introduction to Bayesian Network and Influence Diagram", [Online]:http://people.cs.aau.dk/~uk/papers/pgm-book-I-05.pdf

[33]. Kjærulff,U.,&Madsen,L., (2013). "Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis", *Information Science and Statistics,* 22

[34]. Darwiche,A., (2009). "*Modeling and Reasoning with Bayesian Networks*", Cambridge University Press

[35]. Heckerman,D.Geiger,D., &Chickering,D., (1995). "Learning Bayesian networks: The combination of knowledge and statistical data", *Machine Learning* , 20 (3), 197-243

[36]. Heckerman,D., (2008). "A Tutorial on Learning with Bayesian Networks", *Innovations in Bayesian Networks*,SCI 156, 33-82, Springer-Verlag

[37]. Laurizen, S., (1995). "The EM algorithm for graphical association models with missing data", *Computational statistics and data analysis*,19(2), 191-201

[38]. Myllymäki,P., Silander,T., Tirri,H., &Uronen, P., (2002). "B-course: a web-based tool for bayesian and causal data analysis", *IntJ Artif IntellTools*,11(3), 369-387

[39]. Fenton,N., Neil,M., (2012).*RiskAssessment and Decision Analysis with Bayesian Networks*, CRC Press

[40]. Maglogiannis,I., Zafiropoulos,E., Platis,A., & Lambrinoudakis,C., (2006). "Risk analysis of a patient monitoring system using Bayesian Network modeling", *J Biomed. Informatics*,39(6), 637–647

[41]. Liu,K., Lu,C-F., Chen,C-W., Shen,Y-S., (2012) "Applying Bayesian belief networks to health risk assessment", *Stoch Environ Res Risk Assess*, 26(3), 451–465

[42]. Dimitrova,L., Petkova,K., (2013). "Modeling Injury Risk for Pre-Elite Rhythmic Gymnasts: Bayesian Network Approach",*EIIC*2(1), www.eiic.cz/archive

[43]. Dimitrova,L., Petkova,K., (2014)."A Bayesian Network Application for Estimating the Injury Risk for Pre-Elite Rhythmic Gymnasts", *Informatics & IT today*, 2 (1)

[44]. Dimitrova,L., Petkova,K., (2014). „Bayesian Network-Based Causal Analysis of Injury Risk in Elite Rhythmic Gymnastics", *IJCSSE*, 2(1), 50-61

[45]. Dimitrova,L., Petkova,K., (2014). "Modeling Injury Rate in Women Artistic Gymnastics with a Bayesian Network", *Proceedings in global virtual conference,* 2(1), http://www.gv-conference.com/archive/

[46]. Bayesian Networks and Simulation Software for Risk Analysis and Decision Support,http://www.agenarisk.com/

Corresponding author: Lyudmila Dimitrova
Institution: University "Prof. As. Zlatarov", Burgas, Bulgaria
E-mail: lyudim@gmail.com