SAFETY SCIENCE Monitor

Issue 2 2010

Article 4 VOL 14

THREATS AND VIOLENCE AS A PRECURSOR TO OCCUPATIONAL INJURY: TEXT-MINING OF INSURANCE-BASED INFORMATION ON POLICE OFFICERS AND SECURITY GUARDS IN SWEDEN 2004-2007

TORE J LARSSON
Royal Institute of Technology, School of Technology and Health, SE-100 44, Stockholm, Sweden

KEREM TEZIC
AFA Swedish Labour Market Insurances, Klara Södra Kyrkogata 18, SE-106 27 Stockholm, Sweden

CECILIA OLDERTZ
AFA Swedish Labour Market Insurances, Klara Södra Kyrkogata 18, SE-106 27 Stockholm, Sweden

ABSTRACT

The full text of all occupational injury claims associated with threats or violence from Police Officers and Security Guards reported to the Swedish National workers' compensation insurance 2004 –2007 was analysed with the help of text-mining software. The analysis generated clusters of details on hazardous exposures and accident processes, and the level of information in the clusters describing the different scenarios identified possible practical modifications in training, technology and professional operating procedures. The analysis of the textual information in the reporting of traumatic injuries promises to better represent the dynamics and detail of exposure and accident process and, when applied to valid and representative insurance data with the help of a suitable text-mining software, will provide industry groups and local companies with detailed decision support for safety management and prevention.

Keywords: Text-mining, free text, occupational injury claims, police officers, security guards, threats, violence, Sweden.

1. INTRODUCTION

1.1 Background

In Sweden, although injuries due to occupational violence have decreased during the last 10 years, it is still the second most common precursor to traumatic injuries in the workplace (Larsson et al, 2009). The association between violence at work and fatality is well documented in the U.S. literature (Baker et al, 1982, Bell, 1991, Cone et al, 1991, Toscano and Windau, 1994) and the link is particularly clear for occupational groups carrying arms.

Several questions in relation to the prevention of the occupationally related violence among police officers and security guards would be of interest. What type of typical accident processes are facing police officers? Are the risks age-dependent? Are female police officers at a higher risk of injury compared to their male colleagues? Does the risk depend on the characteristics of the duty? Does the severity of the injury differ between accident processes?

There is little empirical research focusing specifically on the relation between severity of the injury due to violence against police officers and different accident processes leading to traumatic injury. Some studies have analysed the potential dangerousness of domestic disturbance calls (Grennan, 1987, Uchida et al., 1987, Ellis et al., 1993). Rabe-Hemp and Schuck, (2007) analysed violence against police officers focusing on gender differences in assaults on officers, as well as situational and individual factors related to the assault. Their findings suggest that compared to male officers, female officers are at a greater risk of being assaulted in family conflict situations. These results contradict the hypothesis that the need to assert authority may increase the likelihood of violence (Toch, 1992), since having more female officers are thought and found to influence how officers interact with members of the public (Steffensmeier, 1979, Van Wormer, 1981).

In this paper, we analyze the full text of occupational injury claims of police officers and security guards in the context of threat and violence leading to traumatic injury. Both occupational groups are analysed separately in terms of risk exposure, gender and age. The analysis is based on the total Swedish National workers' compensation insurance material (100% coverage of the occupational groups) and the full free-text of the accident process, as depicted on the claims form, by using text-mining analysis.

1.2 Information about occupational trauma

Information about causes of injuries – risk exposures and accident processes at work – is collected as a part of AFA, the Swedish Labour Market Insurances' claims management. The data bank, established in 1988, enables AFA, who provides workers' compensation insurance to 93% of the Swedish labour market, to provide a source of information and a basis for setting priorities for systematic injury prevention in Swedish work life (Larsson 1990, Bråfelt & Larsson 1993, Wännström & Larsson 1996). The risk information facilitates the identification of occupational groups, tasks, activities and machines particularly associated with severe injuries (Larsson et al 2009).

The specific information on the exposure and accident process is based on the answers to three questions on the AFA Insurance claims form related to "activity prior to accident", "accident mechanism" and "contact event". Initially, this text information was coded in an alliterative, telegrammatic form (Heidenstrom 1985). Since 2004, the full free text of the accident process is saved and avaliable for analysis, and from 2007 with the help of text-mining software (Brooks 2007; 2008).

Annually, around 3000 claims for occupational injury due to exposure to threats or violence are filed with AFA (Larsson et al 2009). The distribution over occupations in the period 2005 through 2007 (36 months) is presented in Table 1.

Table 1 Reported occupational injuries due to threat or violence in different occupational groups. AFA Swedish Labour Market Insurances accepted claims 2005-2007.

Occupation	N of cases of threats, violence	Prop. of all claims in the occupation	N of cases per 1000 employed in the occupation	Proportion with >30 days disability or medical impairment
Police Officers	**686**	**39,4 %**	**43,8**	**23,5%**
Security Guards	**505**	**48,7 %**	**33,5**	**51,1%**
Psychologists, Social Workers	**321**	**40,0 %**	**19,2**	**44,9%**
Prison Staff	**120**	**38,6 %**	**16,3**	**29,2%**
Bank Clerks	**430**	**70,1 %**	**16,2**	**93,5%**
Carers, Keepers in psychiatric care	**810**	**54,8 %**	**13,1**	**23,1%**
Enrolled Nurses	**796**	**23,3 %**	**5,0**	**24,1%**

Childcare Staff	**426**	**25,7 %**	**4,5**	**16,0%**
Professional Drivers	**421**	**8,1 %**	**4,5**	**61,3%**
Nursing Assistants	**684**	**28,2 %**	**4,3**	**18,1%**
Preschool Teachers, Recreation Instructors	**325**	**21,3 %**	**4,1**	**11,7%**
Shop Assistants	**809**	**25,6 %**	**4,0**	**84,4%**
Teachers and School Staff	**487**	**18,9 %**	**2,4**	**16,6%**
Military Staff (professional)	**27**	**5,0 %**	**2,4**	**18,5%**
Nurses and Midwives	**203**	**16,1 %**	**1,9**	**39,9%**
Healthcare Specialists	**78**	**15,4 %**	**1,8**	**30,8%**
Storage and Transport Assistants	**242**	**7,6 %**	**1,2**	**40,9%**
Hotel, Restaurant and Kitchen Staff	**75**		**0,7**	**69,3%**
Physiotherapists, OTs, Dental Hygienists	**23**	**5,3 %**	**0,7**	**34,8%**
Administrative Staff	**97**	**5,1 %**	**0,2**	**43,3%**
Other Occupations	**682**		**-**	
Total	**8247**		**2,1**	**39,2%**

This type of occupational injury represents the majority of occupational injury claims from Bank Clerks and Carers, Keepers in psychiatric care. This injury type also represents the largest proportion of severe claims, with extended periods of absence from work and/or assessed permanent impairment, among the claims from Bank Clerks, Shop Assistants, Hotel, Restaurant and Kitchen Staff and Professional Drivers.

The character of violence differs between different occupational exposures. Described in general terms, Bank Clerks and Shop Assistants are exposed to thefts and robberies, Teachers and Childcare Staff have conflicts with children and parents, Service Staff have conflicts with customers, Medical Staff have conflicts with patients and Police Officers and Security Guards are exposed to criminal suspects and offenders.

In the following, all claims associated with threat or violence in the period 2004 through 2007 (48 months) from Police Officers (n=777) and Security Guards (n=558) are treated to a text-mining analysis in order to establish the exposure patterns associated with occupational injury in the two occupational groups. This is the whole population of police officers and security guards who have filed claims for injury due to threat or violence between 2004 and 2007. The free-text of the accident process is digitalised and available for analysis from 2004, and all injuries which have been reported to have occurred in 2004, 2005, 2006 and 2007 have been included.

2. METHODS AND PRINCIPLES OF TEXT-MINING

In order to understand the complexities of processes leading to traumatic injury, free-text descriptions are by far superior to aggregated coded information. Extensive descriptions of accident processes will, when systematically analysed, yield information about typical exposure situations and the dynamic interaction between man, equipment and environment. This often represents “new” knowledge, or rather, knowledge new to those removed from the hazardous exposure (Larsson et al 2008). Stout (1998) has showed the benefits of using information from free-text data not present in the coding system and how the use of this information resulted in improved claims management in the occupational injury insurance domain. Kolyshkina and Van Rooyen (2006) demonstrated how textual data can be included in the claims analysis and be used to improve the prediction of the pay-out value of insurance claims. Brooks (2008) used ten years of Australian workers’ compensation data on the injuries of wood industry workers and delivery drivers to develop a more detailed understanding of hazardous exposures and accident processes with the help of text mining. The methodology of text mining of extensive occupational injury data is well described this work (Brooks 2008; pp 6-18).

One of the main applications of text-mining is to group the documents into meaningful clusters. In order to do that, the documents must be converted into vectors of lower dimensional space. For that, the text-mining software uses the method of Latent Semantic Indexing (LSI). The main assumption is that there is an inherent latent semantic structure in the data that is partially obscured by the randomness of word choice (Deerwester et al.,

1990). In this method, the document collection is numerically represented as matrix of terms by documents. This matrix is then decomposed to derive a latent semantic model. This is done by decomposing the original documents by term matrix into the product of three other matrices of a special form. The process is called Singular Value Decomposition (SVD). These matrices contain singular vectors and singular values, and they show a breakdown of the original matrices into linearly independent components. In the reduced model, all the term-term, document-document and term-document similarities are approximated by values on this smaller number of dimensions. Each term and document is represented by its vector of factor values. Many of these components are very small, and may be ignored leading to a model with fewer dimensions and less noise. The dot product or cosine between vectors representing two objects corresponds to their estimated similarity (Deerwester et al., 1990). It is important to keep the cognitive content of the document by avoiding very low or extremely high number of dimensions and to be guided by what reproduces the greatest amount of variance in the original matrix. In order to increase the performance of SVD, we have ignored non-informative parts of speech such as auxiliaries, conjunctions, prepositions, pronouns. Synonym lists and specified data sets, that contain words which are not significant or meaningful to the specific analysis, have been produced.

Clustering is an analysis technique in which no assumptions are made concerning the number of clusters or cluster structure. Internal consistency and cognitively acceptable content is judged after several iterations have been performed. A good clustering should group together similar objects and separate dissimilar ones. For this, the similarity function that was mentioned above is used by the clustering algorithms. Each cluster is considered in relation to whether it is poorly differentiated or not and whether it seems to overlap with another cluster. In this analysis we used the expectation-maximization (EM) clustering algorithm. The EM algorithm is a general purpose framework for estimating the parameters of distribution in the presence of hidden variables in observable data. The underlying assumption is that the objects to be clustered are drawn from k distributions and the goal is to identify the parameters of each that would allow the calculation of probability $P(C_i \mid x)$ of the given object's belonging to the cluster C_i (Feldman and Sanger, 2007).

The following procedures are performed by EM algorithm:
The initial parameters of k distributions are selected randomly (initialisation). The $P(C_i \mid x)$ for all objects x are computed by using the current parameters of the distributions. All objects are relabelled according to the computed probabilities (E-step). The parameters of the distributions are re-estimated to maximize the likelihood of the objects' assuming their current labelling (M-step). Stopping the iteration process when the change in the likelihood function becomes small (convergence).

The dimension reduction technique with Latent Semantic Indexing maps the enormous documents-terms matrix dimensional space onto a lower dimensional space by representing major associative patterns between documents and terms while getting rid of the noise. Then, by using cluster analysis, the documents are grouped into meaningful clusters identified as typical accident processes. This allows us to focus on risk exposures for different demographic variables and test whether categories differ from each other significantly.

3. OCCUPATIONAL INJURY ASSOCIATED WITH THREAT OR VIOLENCE

3.1 Police Officers' injuries associated with threat or violence

All text reported by the 777 claimants as responses to the queries of the report form ("What were you doing?", "What went wrong?", "How did you sustain the injury?") was included in the analysis. Ninety-five percent of the cases could be successfully allocated to clusters describing exposure situations and accident processes.

A number of questions in relation to gender, age and severity of injury can be posed, ie:
- are gender groups differently exposed to risk of injury from threat or violence?
- are age groups differently exposed to risk of injury from threat or violence?
- does severity of the injury differ considerably between clusters?

Table 2 Clusters of reported occupational injuries due to threat or violence among male (♂) and female (♀) police officers, proportions of severe injuries (>30 days and/or permanent impairment). Swedish Labour Market Insurances accepted claims 2004-2007.

Cluster	Descriptive Terms	N ♂	N ♀	Severe* ♂ (%)	Severe ♀ (%)
1 Taking into custody, transporting to police station, handcuff, resisting person	Put, cell, resist, handcuffs, hold, stand, run, pull, take into custody, struggle, floor, squeeze, backseat, police car, transport, rowdy	53	33	15.1	18.2
2 Spit, saliva; face, mouth, eyes	Saliva, infection, spit, hepatatis, face, mouth, eye, transport, question, patient, police station	30	18	23.3	5.6
3 Car chase, car theft, patrol car	Steal, car, driver, drive, reverse, stab, knife, throat, control, stop, talk, patrol car, jump	21	9	38.1	22.2
4 Police raid, house search, gun, knife, pepper spray	Unit, police raid shoot, house, shot, search, fire weapon, life, pull, service weapon, die, danger, situation	37	8	70.3	62.5
5 Mentally deranged female, clawing, biting, kicking, pulling, hitting (psych. causes)	Female, psychiatric illness, assistance, bite, claw, nail, spit, hepatitis, attack, backseat, take into care, refuse arrest, pull	20	25	10.0	24.0
6 Verbal threat	Threat, earpiece, phone call, note, letterbox, serious, death threat, call, family, worry, kill, experience, menacing, show gun, fear	16	11	81.3	72.7
7 Intervene, stop fight, attacked by violent person	Street patrol, fight, gang, attack, fist, lie, head, neck, back	50	17	22.0	29.4
8 Hit/kicked by fist, knee, foot	Rowdy person, hit, fist, kick, intervention, stop, stand, queueing, arrest, beating	38	10	21.1	20.0
9 Stones/objects thrown (sports/ demonstration)	Supporter, throw bottle, stone, ice hockey, hooligan, bruise, iron bar, shinguards	22	7	9.1	57.1
10 Violent drunk person	Drunk, drunk driver, take into custody, arrest, kick, transport, resistance, knee, cell, elbow, face	69	42	13.0	16.7
11 Violent drug addict	Drug addict, violent resistance, needle, fight, push, tear, wriggle, escape, fall	61	15	22.9	6.7
12 Violent suspect with knife	Violent suspect, knife, punch, unknown perpetrator, resistance, violence, bleed	100	29	22.9	20.6
Clear	Total 741 (95.4%)	517	224		
Unclear	n=36 (4.6%)				
* Severe means >30 days disability or medical impairment					

There are twice as many male compared to female police officers in the injury material. The official statistics from Statistics Sweden (SCB) indicate that the proportions of males and females in the police force are about 0.76 and 0.24 respectively, which is 3:1, and this makes male officers underrepresented and female officers slightly overrepresented in the material. The proportion of age groups for -25, 26-35, 36-45, 46-55 and 56+ years old are 0.018, 0,231, 0.271, 0.275 and 0.205 respectively. The test whether these proportions hold for each cluster will show if there is an over- or under-representation of each age group per cluster.

Based on an analysis of the two-way contingency tables, (cluster by gender, cluster by age and cluster by severity) the null hypothesis that there is no difference of gender related to the exposure leading to traumatic injury, in other words that the two populations (female and male) are the same, has been rejected (Table A.1, column 2) . Our findings are in accordance with those of Rabe-Hemp and Schuck (2007) and point to the fact that gender is an important risk factor. On the other hand, the results indicate that the distribution of age does not differentiate significantly between clusters (Table A.1, column 3). When it comes to the severity of the injury, the null hypothesis, that injury severity is similarly distributed over clusters, is rejected (Table A.1, column 4).

Table A1 The test of independence of demographic categories with respect to clusters (Police officers).

Chi-square test	*Gender (male, female)*	*Age groups (18-25, 26-35, 36-45, 46-55, 56+)*	*Severity of the injury (<31 days disability, >30 days disability or medical impairment)*
Clusters	$\chi^2(11,741) = 36.27^{***}$,	$\chi^2(44,741) = 53.82$,	$\chi^2(11,741) = 100,71^{***}$,
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*			

Table A2 The test of over/under representation of male and female officers for each cluster

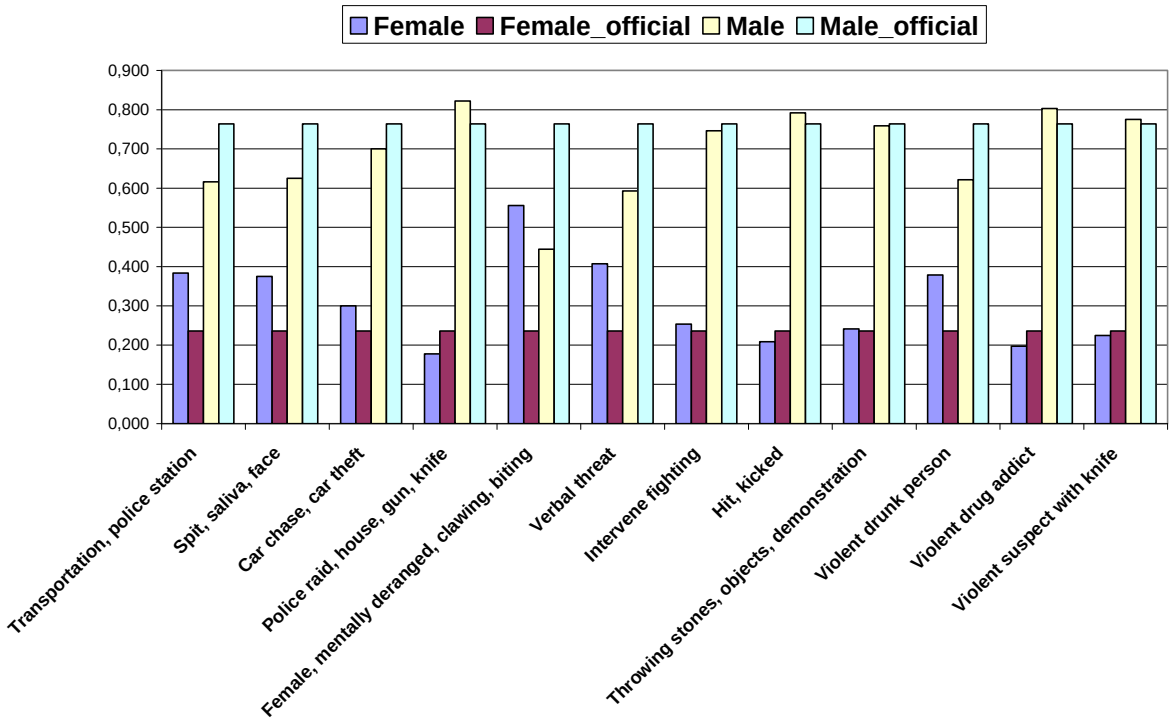
Chi-square test	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*	*Cluster 5*	*Cluster 6*	*Cluster 7*	*Cluster 8*	*Cluster 9*	*Cluster 10*	*Cluster 11*	*Cluster 12*
H_0: p_m =0.76, p_f =0.24	9.73***	4.79**	0.59	0.96	24.57***	4.15**	0.07	2.37	0.0003	11.65***	0.75	0.163
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*												

Table A3 The test of over/under representation of age categories for each cluster (Police officers)

Chi-square test	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*	*Cluster 5*	*Cluster 6*	*Cluster 7*	*Cluster 8*	*Cluster 9*	*Cluster 10*	*Cluster 11*
H_0 :p_{-25} =0.018, p_{26-35} =0.231 p_{36-45} =0.271, p_{46-55} =0.275 p_{56+} =0.205	42.71***	75.89***	20.50***	24.43***	75.55***	12.06**	31.32***	30.27***	32.55***	93.71***	63.19***
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*											

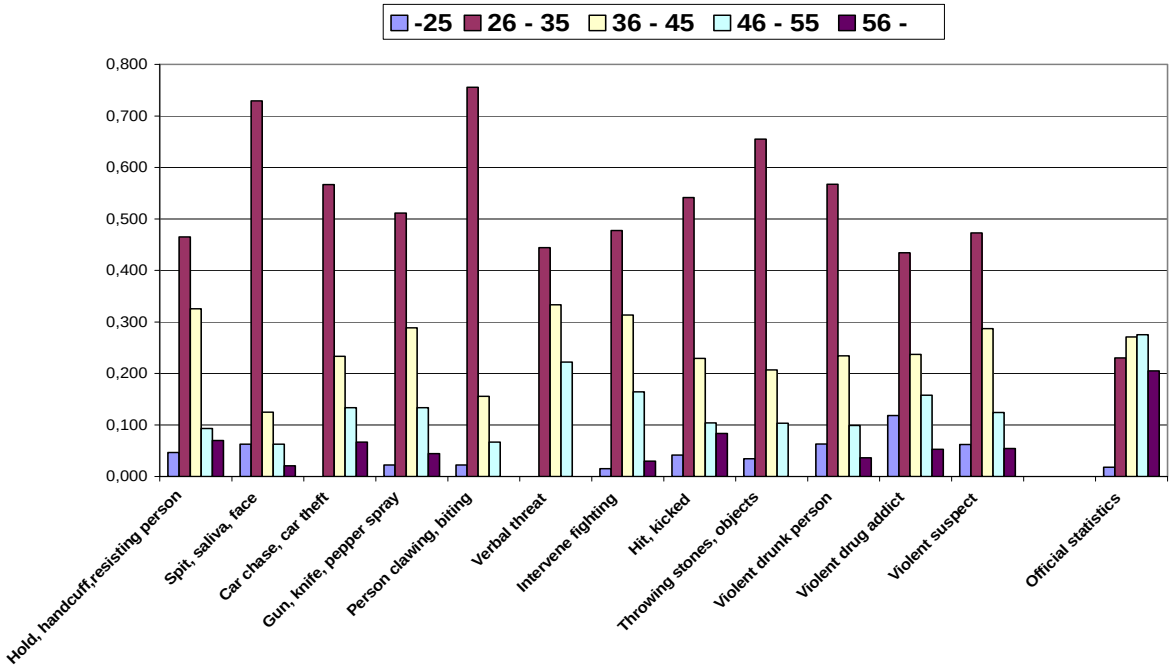
In terms of organisational exposure, the chi-square tests of the null hypothesis that for each cluster the proportion of males and females are about 0.76 and 0.24 respectively, have been rejected for the first, second, fifth, sixth and tenth clusters. In order to visualise the situation, Graph 1 below, shows the distribution of gender by cluster. For each cluster, the second and fourth staples (named female-official and male official), correspond to the proportion of female and male police officers within the police force in Sweden (SCB, Statistics Sweden). In all the clusters mentioned above, female police officers are overrepresented and thus over-exposed to injury risk. Female officers are strongly overrepresented in cluster 5 (person clawing, biting, kicking, pulling, hitting; psychiatric causes) – 56% against 24% females in the force. The same is true for cluster 6 (verbal threat) - 41% of the claims were from female officers.

Graph 1



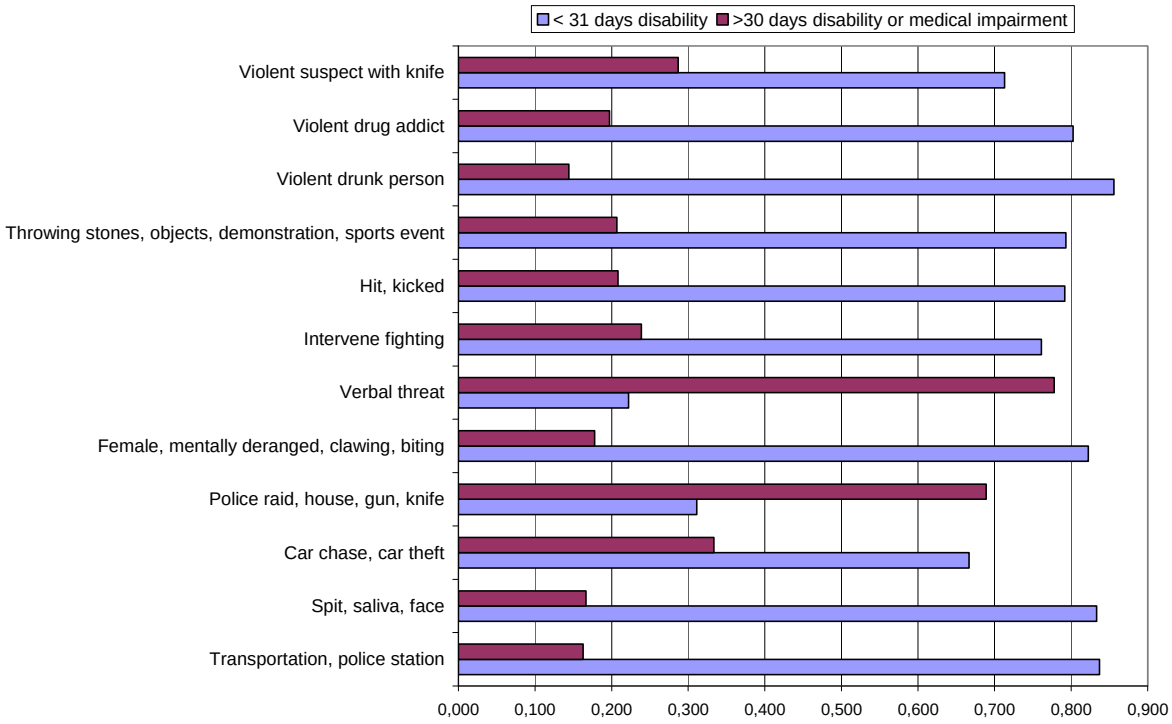
When it comes to age, the chi-square tests of the null hypothesis that for each cluster the proportion of age groups for -25, 26-35, 36-45, 46-55 and 56+ years old are 0.018, 0,231, 0.271, 0.275 and 0.205 respectively is rejected for all the clusters (Table A.3). All clusters show an overrepresentation of young Police Officers (18-35) of between 2 to 3 times the distribution in the occupation. In cluster 11, Officers aged 18-25 are 6.6 times more common than their representation in the force (Graph 2). These results point to the fact that regardless of the operational procedure, young police officers are more exposed to injury risk.

Graph 2



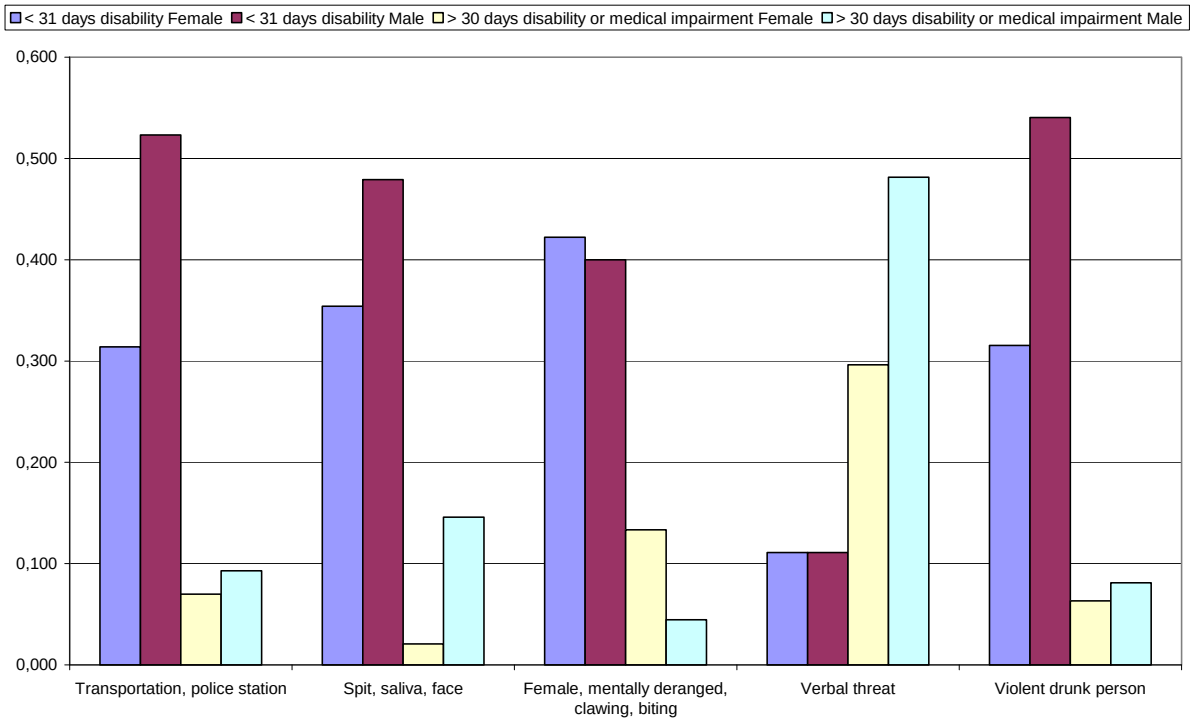
The exposure which includes the highest proportion of severe claims is cluster 6 (verbal threat), where 4 out of 5 claims are associated with long term absence from work (Graph 3).

Graph 3



For the clusters where the over/under representation of gender per cluster is significant (clusters 1, 2, 5, 6 and 10), one can see that in in cluster 2 (spit, saliva, face, mouth eyes) and in cluster 6 (verbal threat), male officers are more likely than female officers to suffer severe consequences. On the other hand, in cluster 5 (mentally deranged female), female officers have a much higher proportion of severe injuries than their male colleagues (Graph 4).

Graph 4



3.2 Security Guards' injuries associated with threat or violence

All text reported by the 558 claimants as responses to the queries of the report form ("What were you doing?", "What went wrong?", "How did you sustain the injury?") was included in the analysis. Ninety-four percent of the cases could be successfully allocated to clusters describing exposure situations and accident processes.

Table 3 Clusters of reported occupational injuries due to threat or violence among male (♂) and female (♀) security guards, proportions of severe injuries (>30 days and/or permanent impairment). Swedish Labour Market Insurances accepted claims 2004-2007.

CLUSTER	*DESCRIPTIVE TERMS*	*N♂ ♂*	*N ♀♀*	*Severe* ♂ (%)*	*Severe ♀ (%*
1 Patrolling pub, restaurant	*Patron, restaurant, customer, patrolman, drunk, removal, turn away, saliva, head-butt, beating*	*50*	*12*	*46.0*	*41.7*
2 Car-park,traffic warden	*Car park attendant, parking fine, vehicle, driver, colleague, parking, friend, knocked down*	*42*	*23*	*40.5*	*47.8*
3 Shoplifter,take into custody	*Shop, shop-walker, take into custody, male shoplifter, steal, plainclothes security guard, violent resistance, bump, suffering*	*47*	*23*	*40.4*	*52.1*
4 Robbery	*Robbery, ATM, bank, value, masked, knock down, automatic weapon*	*33*	*8*	*87.8*	*75.0*
5 Subway; theft, shoplifting	*Assault alarm, subway, theft, shoplifting, take into custody, pepper-spray, keeping order, disturbance, resistance, injury, guarding, intervention*	*75*	*24*	*37.3*	*33.4*
6 Ticket control	*Ticket control, entrance, ticket, bus, subway, assault, passenger, colleague, try, control*	*15*	*3*	*26.7*	*-*
7 Armoured truck robbery	*Robbery of armoured truck, money bag, robbery, servicebox, armed robbery, transport, armed, vehicle, order, yell, ATM*	*56*	*13*	*87.5*	*100*
8 Patrolling shopping centre	*Patrolling guard, shopping centre, plainclothes guard, violent resistance, guard, large spanner*	*24*	*9*	*50.0*	*66.7*
9 Attack by addicted or aggressive person	*Assault, public transport, black eye, addict, rebuke, centrum, disturbance, concussion*	*58*	*12*	*51.7*	*50.0*
Clear	*Total 527 (94.4%)*	*400*	*127*		
Unclear	*n=31 (5.6%)*				
** Severe means >30 days sick or medical impairment*					

There is an overrepresentation of female security guards only in clusters 2 and 3, probably reflecting the actual female overrepresentation among staff employed as traffic wardens and shop-walkers.

In order to test if there are systematic differences between exposure situations in terms of gender, age and severity of the injury, the same chi-square test as in Section 3.1 was performed on the Security Guards material.

The results from chi-square test (Table A.4, column 2) indicate that the null hypothesis - no difference of gender due to accident process leading to traumatic injury - has not been rejected. The result is in contrast to what we found for Police Officers. Furthermore, there is no difference between age groups in terms of type of risk exposure (Table A.4 , column 3).

Table A4 The test of independence of demographic categories with respect to clusters (Security guards).

Chi-square test	*Gender (male, female)*	*Age groups (18-25, 26-35, 36-45, 46-55, 56+)*	*Severity of the injury (<31 days disability, >30 days disability or medical impairment)*
Clusters	$\chi^2(8,527) = 12.32,$ $p = 0.14$	$\chi^2(32,527) = 24.83,$ $p = 0.81$	$\chi^2(8,527) = 78{,}99^{***},$ $p < 0.0001$
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*			

Table A5 The test of over/under representation of male and female security guards for each cluster

Chi-square test	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*	*Cluster 5*	*Cluster 6*	*Cluster 7*	*Cluster 8*	*Cluster 9*
H_0 : p_m =0.72, p_f =0.28	2.30	1.75	0.81	1.46	0.69	1.14	2.88	0.61	4.09**
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*									

Table A6 The test of over/under representation of age categories for each cluster (Security guards)

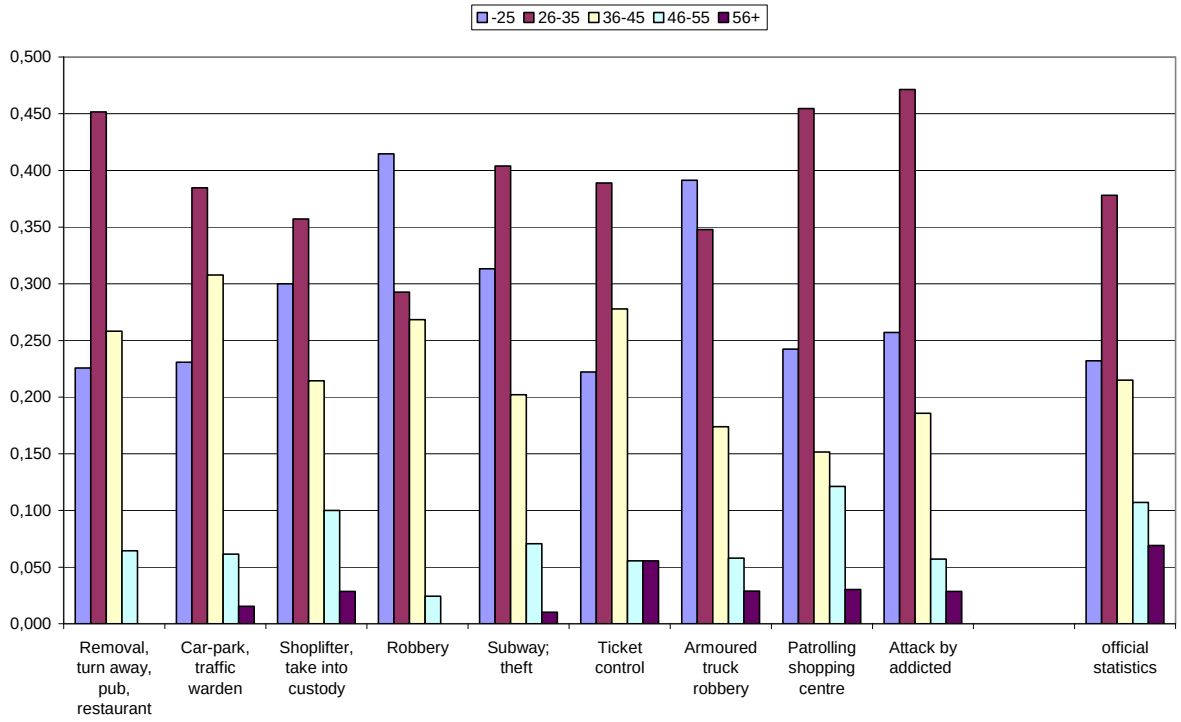
Chi-square test	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*	*Cluster 5*	*Cluster 6*	*Cluster 7*	*Cluster 8*	*Cluster 9*
H_0 : p_{-25} =0.232 p_{26-35} =0.378 p_{36-45} =0.215 p_{46-55} =0.107 p_{56+} =0.069	6.89	7.11	3.37	13.61***	9.61**	0.94	11.77**	1.82	5.45
significant at 10 percent, ** significant at 5 percent, * significant at 1 percent*									

Common to both genders is that the highest proportions of severe injuries are associated with armed bank robberies and the robberies of armoured trucks. The null hypothesis, that the injury severity categories belong to the same population with respect to exposure, is rejected (Table A.4, column 4).

The results for over/under representation of gender groups indicate that the hypothesised proportion of male and female security guards (about 0.72 and 0.28 respectively) is rejected at α =0.05 level for only the last cluster, attack by addicted, where female security guards are overly represented (Table A.5).

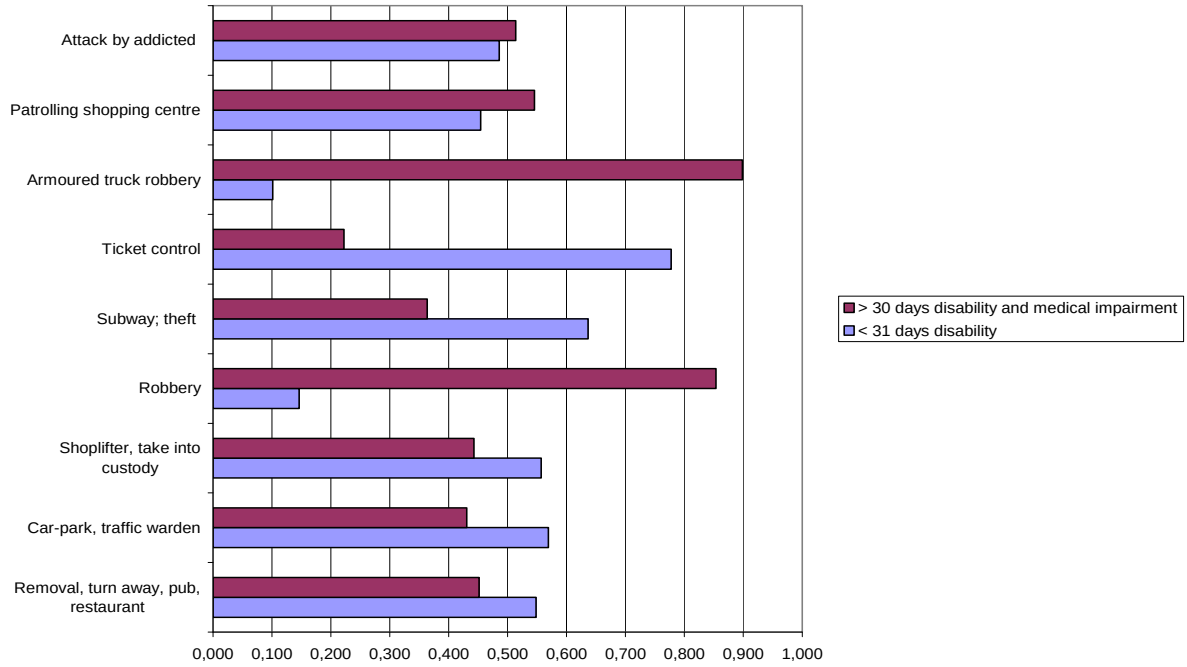
The null hypothesis - that the age distribution in the clusters is the same as in the profession - is rejected at α =0.05 level for the forth (robbery), fifth (subway theft) and seventh (armoured truck robbery) clusters (Table A.6). The over/under representation of age groups is illustrated in Graph 5, where the last staples correspond to the age distribution provided by Statistics Sweden. It is clear that the youngest age group (-25) is overly exposed to injury risks for the three clusters mentioned.

Graph 5



Although police officers and security guards have different type of duties, the association between the severity of the injury and accident process display similarities: the exposures related to direct confrontation with armed resistance or vengeance related threat/violence are closely associated with more than 30 days disability or medical impairment (Graph 6).

Graph 6



4. CONCLUSIONS AND IMPLICATIONS FOR PREVENTION

This analysis of occupational injury claims from Police Officers and Security Guards related to threats and violence is undertaken with a different methodology compared to earlier studies. The text analysis software tools represent new possibilities for exploratory, yet systematic, insights into the particulars and details of hazardous exposures described and included in the large amounts of data recorded in the National workers' compensation system. The analysis helps to sort exposure situations and accident processes into cognitively distinct and meaningful clusters by assuming that there is a latent semantic structure in the data. By linking demographic variables to the resulting exposure information, detailed suggestions for safety management and injury prevention can be formulated.

The Swedish workers' compensation scheme represents National coverage, ie all Police Officers and Security Guards in the country are covered and included. There is no reason to believe that the system suffers from systematic under-reporting or missing data. The actual age/gender distribution in everyday operations in these professions, however, is unknown and this must be considered when drawing conclusions from the presented data.

Ninety-five percent of all accident processes could be acceptably and credibly allocated to clusters possible to differentiate in relation to exposure. The analysis of the text information of the claims forms generates detailed descriptions of occupationally and situationally specific exposures suggesting a typology of such dynamic (accident) scenarios which lead to injury. The level of information in the clusters describing the different scenarios is such that measures of risk management and control could be developed through training, operational routines, or technical resources, i.e.:

further differentiate operational procedures/techniques among police officers between arrests of violent suspects, mentally deranged suspects, drunk suspects, addicted suspects;

develop further the operations of police officers manually taking suspects into custody according to the different environment - suspect's car, backseat of patrol car, cell;

carefully analyse the operational procedures in terms of using female police officers for certain types of interventions;

clarify the role and response, and conduct preparatory training, of the security guards likely to be exposed to armed robbery.

Acknowledgements: We have benefited from the comments of the participants at presentations in SAS Global Forum in Washington DC and SAS Analytics Conference in Copenhagen 2008-2009.

REFERENCES

Baker, S. P., Teret. S. P. and Dietz, P. E. (1980). Five year risk of assault on employees in a psychiatric hospital, Safety Science, 18, 113-124.

Bell, C. A. (1991). Female homicides in United States workplaces, 1980-1985, American Journal of Public Health, 81, 729-732

Brooks, B (2007). Know-Risk: design of an architecture for an acute occupational injury risk decision support system. PhD thesis. Monash University Accident Research Centre. Melbourne.

Brooks, B (2008). Shifting of focus of strategic occupational injury prevention; Mining free-text, workers' compensation claims data. Safety Science, 46 , pp 1-21.

Bråfeldt, O & Larsson, TJ (1993). Sanning och Konsekvens - lokalt skadeförebyggande med hjälp av Trygghetsförsäkringens skadedata. (Truth and Dare - local injury prevention with the help of injury data from the No-Fault Liability Insurance). IPSO Factum 39, Stockholm 1993.

Cone, J. E., Daponte, A., Makofsky, D., Reiter, R., Becker, C., Harrison, R.J., and Balmes,J. (1991). Fatal injuries at work in California. Journal of Occupational Medicine, 33, 813-817.

Deerwester S., Dumais, S. T., Fournas, G. W., Harshman, R. and Landauer, T. K. (1990). Indexing by Latent Semantic Analysis, Journal of the American Society for Information Science, 41, 391-407

Ellis, D., Choi, A., & Blaus, C. (1993) Injuries to police officers attending domestic disturbances: An Empirical Study. Canadian Journal of Criminology, 35, 149-168.

Feldman, R. and Sanger, J. (2007). The Text Mining Handbook, Cambridge University Press.

Grennan, S. A. (1987). Findings on the role of officer gender in violent encounters with citizens. Journal of Police Science and Administration, 15, 78-85.

Heidenström, P: Accident statistics, coding systems and the New Zealand experience. In RA, Protti, DJ & Scherrer, JR (Eds): Role of Informatics in Health Data Coding and Classification Systems. Elsevier, IFIP-IMIA, 1985.

Rabe-Hemp, C. E. and Schuck, A. M. (2007). Violence against police officers. Are female officers at greater risk? Police Quarterly, 10, 411-428

Jansson M, Larsson TJ, Normark M, Oldertz C, Tezic K (2008). Allvarliga arbetsskador och långvarig sjukfrånvaro. AFA Försäkring, Stockholm

Kolyshkina, I. and Van Rooyen, M. (2006). Text Mining for Insurance Claim Cost Prediction, in Lecture Notes in Computer Science, Springer Berlin, 3755, 192-202.

Larsson, TJ (1990). Accident information and priorities for injury prevention (Doctoral Dissertation, Royal Institute of Technology, Stockholm). IPSO Factum 21, Stockholm.

Larsson TJ, Normark M, Oldertz C, Tezic K (2009). Allvarliga arbetsskador och långvarig sjukfrånvaro. AFA Försäkring, Stockholm

Larsson, TJ, Jansson, M & Brooks, B (2008). Text-mining of insurance-based information: Decision support for local safety management. Paper to Workingonsafety.net IV, 30.9 – 3.10 2008, Crete, Greece.

Steffensmeier, D. J. (1979). Sex roles orientation and attitudes toward female police. Police Studies, 2, 39-42

Stout, N. (1998). Analysis of narrative text fields in occupational injury data, in Occupational Injury: Risk, Prevention and Intervention, A.-M. Feyer and A. Williamson. London, Taylor & Francis: 15-20.

Toch, Hans (1992). Violent men: an inquiry into the psychology of violence (4th ed.). Washington DC: American Psychological Association

Toscano, G. and Windau, J. (1994). The changing character of fatal work injuries. Monthly Labor Review, 117, 17-28.

Uchida, C. D., Brooks, L. W. and Koper, C. S (1987). Danger To Police During Domestic Encounters: Assaults On Baltimore County Police, 1984-86, Criminal Justice Policy Review , 2: 357-371

Van Wormer, K. (1981). Are males suited to police work? Police Studies, 3, 41-44.

Wännström, T & Larsson, TJ (1996). Using information properly makes for effective prevention, in Menckel, E & Kullinger, B (Eds): Fifteen years of Occupational Accident Research in Sweden. Swedish Council for Work Life Research, Stockholm.