SAFETY SCIENCE Monitor

Issue 1 2009

Article 1

VOL 13

TEXT-MINING OF INSURANCE-BASED INFORMATION: DECISION SUPPORT FOR LOCAL SAFETY MANAGEMENT

TORE J LARSSON
Royal Institute of Technology

MALIN JANSSON
AFA Swedish Labour Market Insurances

BENJAMIN BROOKS
University of South Australia

Abstract - Based on underlying perspectives, different data tend to be included in reporting systems for occupational accidents and injuries. Information sought mainly for medical, legal, economic or statistical reasons can reduce or misrepresent the information needed for prevention.

Actual detail of accident processes provided by local operators *in verbatim* will serve to counter misconceptions of risk distributions in the organisation and direct the focus to possible prevention measures.

The analysis of free-text reporting of traumatic injuries, accidents and incidents promise to better represent the dynamics and detail of exposure and accident process and, when applied to valid and representative insurance data on occupational injury with the help of suitable text-mining software, will provide industry groups and local companies with decision support for prevention and safety management.

The trial and development of text-mining software for injury prevention, in conjunction with the industry group safety information system, is described and examples from different exposure areas are given.

1 BACKGROUND

1.1 Information about occupational trauma

It is commonly assumed that injuries can be prevented with the help of statistics. A good national occupational accident statistics system in an industrialized country - it is believed - will provide the basis for regulations and recommendations, and these will in turn decrease the number of occupational injuries and diseases in the land (EU-OSHA 2003).

However, looking back along the historic process of work environment improvement it is fair to say that the link between injury statistics and improved working conditions is conjectural. In a country like Sweden, with a reputation for massive work environment reform and several extensive occupational accident information systems, injury statistics have not really been utilized in applied injury prevention.

The traditional output of National information systems on occupational injuries is overview, aggregate statistics. Legally required reporting to a national authority information system is typically suffering from limitations, whether in Sweden or elsewhere. Reporting from the workplace generally means low reliability on injury consequences, thus not providing the information about the *accident process* with valid information on the *injury*. It will also, depending upon the legal structure and the compliance control in the system, be reports filtered through issues of guilt, fault, blame and damages.

Injury information systems have also been established in hospitals in a number of countries. Here, information on the medical trauma associated with the accident, and the ensuing rehabilitation consequences, seems to be more readily at hand. The shortcomings in most of these systems are a low degree of reliability and validity of information on events and conditions preceding injuries, mirroring the bias of the medical professions towards the *injury* rather than the *accident* . This has been limiting the preventive potential of these systems.

It is still uncommon for a hospital-based injury reporting system to enquire about the occupation of the injury victim, thus ignoring information on exposure and epidemiology. Surprisingly few of the hospital systems keep track of long-term and severe consequences of injuries, i e permanent medical disabilities, and thus lack information about the total burden of disability in society due to injuries.

Depending on whether injury information is collected for legal, medical or economic purposes, varying proportions of the actual injury problem is missing - underreporting is as chronic to occupational trauma as it is to crime. To establish a true, statistical representation of the real injury problem in society is therefore complicated (Larsson, 1990). This, however, is not a prerequisite for prevention.

To those who have no personal stakes in the field of occupational health and safety it seems that reductions in occupational injuries and improvements of occupational health problems are brought about by market-defined, competitive technical development, removing humans from dangerous exposures - in many instances removing dangerous jobs altogether.

1.2 Fallacies preventing prevention – attitudes and cultural traits?

A general notion that underlies the perceived necessity of a comprehensive system of accident "statistics" - and a view often, unfortunately, voiced by those who propose more research into accidents - is that there is a chance relation between accident and injury.

This very human notion rests on deeply rooted, generally held attitudes and metaphysical conclusions drawn from the fact that some die and others escape unharmed from, seemingly, similar accidents. This is always true; on the experience level, the time/space configurations always mean that accidents will be perceived as stochastic rather than continuous phenomena. This is probably the core of the definition of an accident.

But it is also true that, on national or industry-wide level, similar exposures, similar operations under similar conditions, tend to produce a very stable volume of similar injuries to those that are exposed. There is nothing metaphysical about this.

A related fallacy is the general, popular attitude all over the world, that individual behavior and "lack of care" is to blame for accidents. This pre-scientific notion is a statement on cultural and ideological climate and seems to permeate all levels of society and makes prevention hard to motivate. If individual behavior and risk-taking would explain why occupational accidents happen, it is strange indeed that cabinet-makers much more often than others miss fingers or that fishermen and professional divers tend to drown at sea.

Different exposures tend to produce different injuries. A low level of physical risk, associated with low probabilities of death and severe medical trauma, tend to produce low grade physical injuries at times when the man-machine system becomes unstable, whereas a high level of physical risk, associated with higher probabilities of death and severe medical trauma, typically is associated with severe physical injuries to its operators. Human behavior and error rates vary, but operators in hazardous jobs tend to be more competent than others in coping with risks (Hovden & Larsson, 1987).

A third fallacy, and an influential and insistent one, is the simplistic economic rationalism of those who propose strict experience rating in workers' compensation systems in order to "send the right signals" to companies. None of the proponents of this approach really know the microeconomics of experience-related premiums; they don't know the reasoning and the multitude of factors relevant to the management of a medium-sized company operating in an area with a high degree of exposure to physical risk to its workers. And the employer who pays the premium is rarely the worker exposed to injury risk.

1.3 To motivate prevention

An injury problem must be perceived as important to be addressed. The importance rests with the severity of the trauma, the loss of life or degree of loss of quality-of-life. Very few in society can quantify this - insurance policies attempt to quantify what cannot be quantified.

A *selective* approach means defining an injury problem at a level, where it can be addressed with prevention. That is, combining information so that a reasonable degree of precision is acquired. For instance, instead of combining industry with age and sex in computing all claims for occupational injuries over one year, you can combine the occupations farmers and agricultural workers with type of external agency for all claims for injuries to fingers, hands and lower arms with a certain level of severity over the last three years (Larsson 1990b).

With a closer look at which external agencies (machines, tools, animals) claim the highest average consequences in injury severity and costs, the compensation data will yield activities and accident mechanisms associated with these exposures (Persson & Larsson 1991; Larsson & Lindqvist 1992; Larsson & Forsblom 2005). This can focus preventive intervention in the form of technical redesign of details of machines, special attention to details of certain tasks for training and retraining efforts, redesign or the introduction of new protective equipment, or, at the bottom of the list, selective information efforts targeted at the groups in question.

The purpose of a detailed and dynamic national overview of work-related injury and disease is, apart from being an excellent tool for the study of occupational injury epidemiology (Larsson 2003), to provide insured companies and exposed workers with a tool for risk analysis, local safety management and applied injury prevention.

2 DECISION SUPPORT SYSTEMS

2.1 Insurance data and information for prevention

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 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 and Larsson 1996). The risk information facilitates the identification of occupational groups, tasks, activities and machines particularly associated with severe injuries (Jansson et al 2008).

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 SAS text-mining software (Brooks 2007).

The Victorian Workcover Authority, Worksafe, the statutory authority administering the workers' compensation scheme in the state of Victoria, Australia, has detailed knowledge of the distribution of occupational risk in the Victorian labour market (Larsson & Field 2002, 2002b, Field et al 2000). Additional information on occupational exposure, provided by way of a merged computer run with data from the Statistics Australia Census, has facilitated detailed analysis of occupationally specific risks (Larsson et al 2001).

Digital compensation claim records for workers in Victoria, Australia has been collated for five 'study periods' by the Victorian WorkCover Authority (VWA) - 1 December 1992 to 31 March 1994, and then the same 16 month periods for years 1994-1996, 1996-1998, 1998-2000 and 2000-2002 (Larsson and Field 2000). These collated records include the text categories 'Injury & Accident Text' and 'Claims Text' were provided to the researcher in a de-identified form. The two text categories were included in the text mining process.

2.2 Risk assessment and decision support

Assessing risks is a qualitative undertaking. It takes the expert opinions of those familiar with the physical and social processes exposing people to hazards. It is done with the help of experience and knowledge rather than with assumed probabilities. It is thus something which lends itself to the structures of a decision support system.

The accident problems are qualitative in nature, solutions require information processed with this in mind; simplistic statistics will not be of any use. The difficulties of local injury prevention can be overcome with the linking of the local field to the national or industrial overview; if tools for local risk assessment are structured in the same way as claims data, comparisons will be possible.

2.3 Language as the optimal code

In order to understand the complexities of processes which lead 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.

Extensive free text information about accident processes, systematically analyzed, will also serve to counteract erroneous conclusions about the generality of human error and point to details and characteristics in the process, which could be amenable to counter measures.

Computer software for sophisticated analyses of free text (e.g. SAS Text Miner) is now available, and will, after adaption to the logical structure of the accident process and exposure arenas, yield detailed information defining and separating clusters of accidents.

3 EXAMPLES OF APPLIED DECISION SUPPORT

3.1 Fractures among Australian Woodworkers

The 2 508 pre-processed claim records for Wood Industry Workers from 1992-2002 in the Victorian Workcover Authority system were clustered using SAS Text Miner. The clusters accounted for 85 per cent of records. 226 records were placed into 'junk-clusters' that had no clear pattern and 5.9% or 149 records remained un-clustered. Most significant are traumas to the fingers from contact with machinery, and back and arm strains from manual handling (Brooks 2007).

Severe injuries should be prioritized for preventive activities as they create a greater burden on workers and the claim system. Deeper analysis of this data reveals systematic occurrences that lead to the injury. Re-clustered text data from the 373 cases of injuries among the Wood Industry Workers coded as "trapped by moving machinery" leading to fracture or amputation are shown in Figure 1.

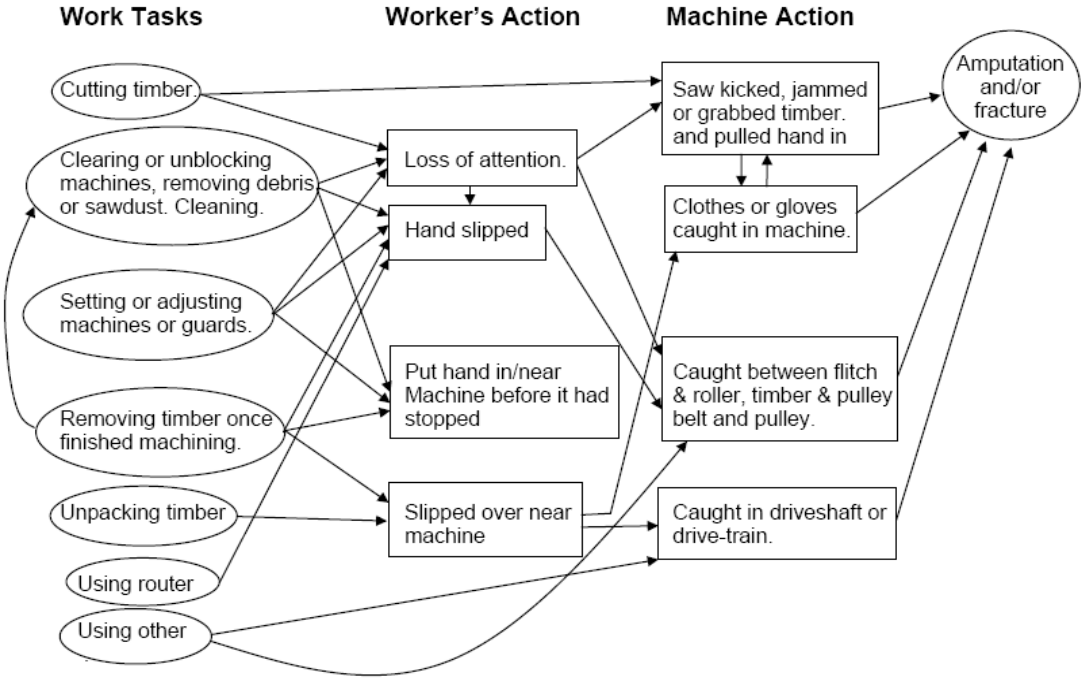


Figure 1 Wood Industry Workers trapped by moving machinery 1992-2002 (n=373) (Brooks 2007)

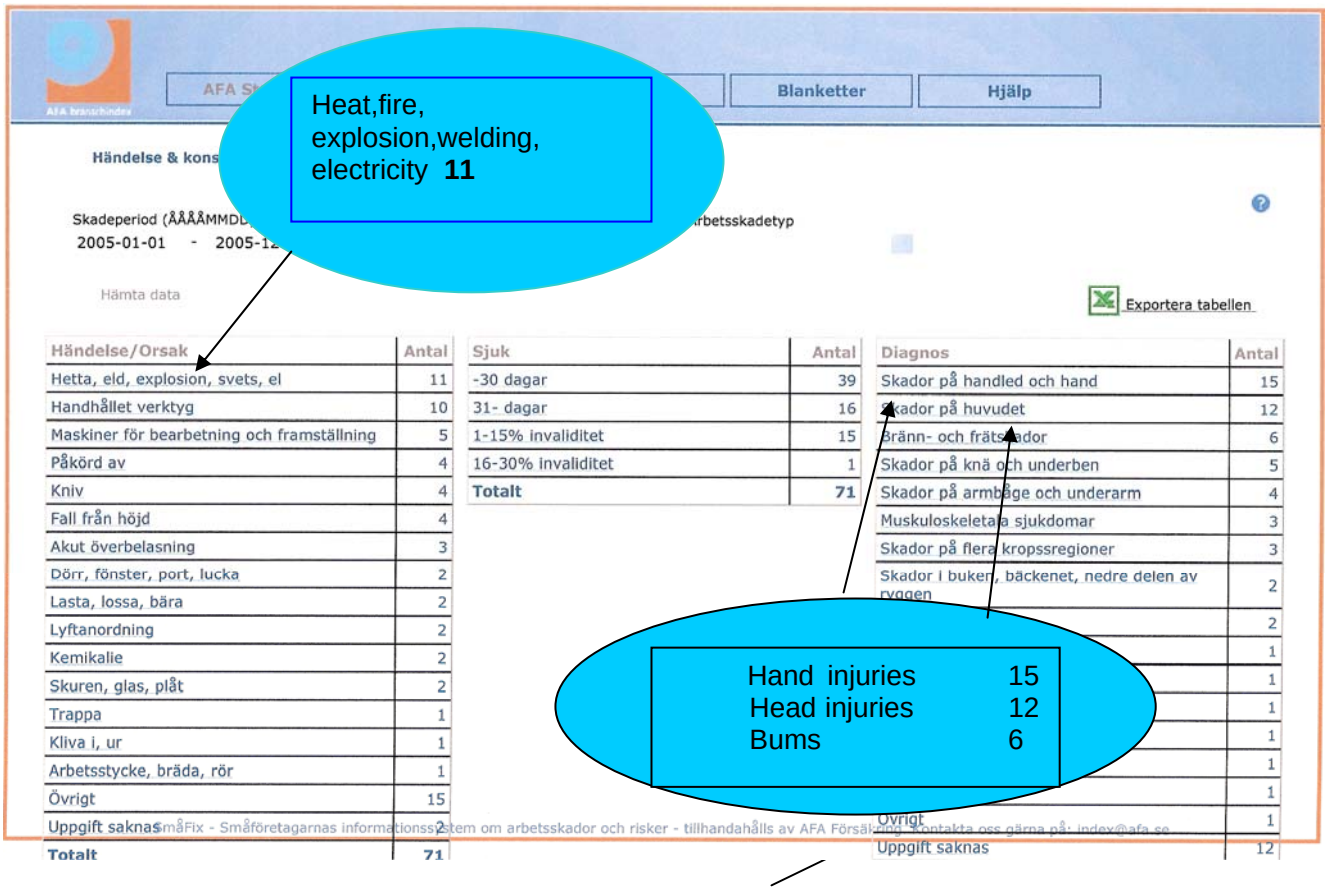
The accident process can be described with the help of a partial bow-tie model (Hale, Goossens et al 2004). The injuries are 87% finger, thumb or hand injuries, 6% forearm and wrist injuries. It is clear that the tasks preceding accident differ in a way which point to quite different prevention measures. The type of machine action (kick-back, jamming, catching glove) and the part of machinery involved (saw, flitch, roller, pulley, belt, driveshaft) will identify different counter measures.

This is an extended and precise form of decision support using the unformatted 'Injury & Accident Text' and 'Claims Text' from the Victorian Workcover Authority claims forms.

3.2 Burn injuries in Swedish car repair shops

When Swedish AFA Insurance claims data is merged with Statistics Sweden's Industry Register, businesses can be identified by type of activity, size of operation and accident processes leading to severe injury described by open text. A computer run on all severe injuries in small car repair shops (employing fewer than 20) in Sweden during 2005 resulted in 71 cases of injury leading to sick leave of 30 days or more and/or medical impairment.

Eleven of the injuries (15%) were the result of heat, fire, explosion, welding or electricity leading to scalds and burns, extended period of sick leave and some percentage of medical impairment. The free-text descriptions of these injuries revealed that three out of the eleven injuries had virtually identical accident processes.



Händelse/Orsak	Antal
Hetta, eld, explosion, svets, el	11
Handhållet verktyg	10
Maskiner för bearbetning och framställning	5
Påkörd av	4
Kniv	4
Fall från höjd	4
Akut överbelasning	3
Dörr, fönster, port, lucka	2
Lasta, lossa, bära	2
Lyftanordning	2
Kemikalie	2
Skuren, glas, plåt	2
Trappa	1
Kliva i, ur	1
Arbetsstycke, bräda, rör	1
Övrigt	15
Uppgift saknas	2
Totalt	**71**

Sjuk	Antal
-30 dagar	39
31- dagar	16
1-15% invaliditet	15
16-30% invaliditet	1
Totalt	**71**

Diagnos	Antal
Skador på handled och hand	15
Skador på huvudet	12
Bränn- och frätskador	6
Skador på knä och underben	5
Skador på armbåge och underarm	4
Muskuloskeletala sjukdomar	3
Skador på flera kroppsregioner	3
Skador i buken, bäckenet, nedre delen av ryggen	2
	2
	1
	1
	1
	1
	1
Övrigt	1
Uppgift saknas	12

Table 1 Car repair shops 2005: 15% scalds and burns

In the three cases, a young worker had short-circuited the spray can when he sprayed the battery, thereby igniting the can, splitting it and starting a severe fire.

> *"The young apprentice burns himself when spray cleaning the engine of a car. He starts a fire in the workshop by splitting the spray can; the staff manages to put out the fire but the car becomes a write-off. The apprentice receives second-degree burns to his hands and arms and he will be off work for at least four months. He will also undergo repeated surgical treatments and, after a few years,be given a permanent degree of impairment of at least 20%." (Larsson 2006)*

There were several other burns and scalds injuries reported among young car repair workers; e.g. hot coolant spraying out from over-heated engine, and battery-acid from exploding battery due to fumes from cleaning agent causing fire.

Suggested prevention measures (as given by the manager of one of the car repair shops involved) were that certain equipment used in the shop for cleaning engine parts and close to the battery should be abolished and specific "hot routines" drawn up and shop supervisors given the time to go through these procedures with staff. Customers were informed of the change of cleaning fluids and the reasons for this were explained.

3.3 Occupational injuries among Swedish police officers

All accepted work-related injuries among Swedish police officers in 2005 and 2006 were subjected to a text-mining analysis in order to understand the details of the main hazardous exposures and accident processes leading to long-term sick leave and medical impairment.

Cluster	Text describing the accident process	Number of documents	Proportion (%) on sick-leave for more than 30 days and/or medically impaired	Proportion medically impaired (%)
1 Training with dog/weapon	+ practice, hay, +knee,fast,+slip,+wrist, +exercise, +dog,+pain,+weapon,+control,+ear,+shoot,+feel,+pain	32	34,4	12,5
2 Fire, inhale	+ fire, + smoke, poisonous, smoke, + garage, evacuate, company, breathing in, + smoulder, + finish, + building, + service, + breathing, pepper spray	16	12,5	6,3
3 Arrest; violence	+violent, + squeeze, + resistance, + injury, + take into custody, + police officer, + put down, do, + a lot, right, + hold, + result, + get, then,+ colleague, + intervention, grip, + blood.	162	26,0	3,7
4 Arrest; psycho	+ mentally, + sick, + head, + threaten, + attack, + shoot, + place, + get, + order, outer, + hit, + throw, + escort, + gun, there.	35	31,4	0,0
5 Arrest; drunk	+ intervention, + kick, + eye, + face, +hit, + kick, + punsch, + suspect, + take into custody, + suspect, + violent, + spit, + mouth, + resistance, + drunk.	280	19,6	3,2
6 Arrest; cut/glass	+ open, + door, + pane, + apartment, + lock, + room, + crush, + cut, + finger, + hand, + car, + suspect, up to, + walk, + pull out.	40	20,0	5,0
7 Bitten	+ dog bite, + dog, + cat, + bite, police dog, + hand, + bite, + attack, + chop, + target figure, + instructor, + control, + jump, + help, + get stuck.	38	21,1	10,5
8 Threat	+ threat, + note, child, family, morning, + address, + letter box, + newspaper, + fetch, worry, + serious, + find, threatening, not, + feel.	10	60,0	0,0

9 Traffic	**+ driver, + drive, + collision, + vehicle, + car, +policecar, + neck, + stop, + go, + patrolcar, + road, + see, + high, + sit, not.**	**69**	**5,8**	**4,3**
10 Phys training	**+ knee, + education, + physical, + play, + turn, + training, indoor hockey, + ball, + opponent, + foot, + crooked, + jump, + joint, + floor, + get stuck.**	**103**	**42,7**	**16,5**
11 Needle	**+ hypodermic, + jab, + plastic bag, + needle, narcotic drug, + syringe, + search. + look, + lie, + scratch, + client, + pocket, + table, warranted search, hepatite.**	**17**	**82,4**	**0,0**
12 Fall	**+ fall, + slip, over, + staircase, + fall, + drop, + step, +ground, + road, + back, +land, + drop, down, + crooked, + run.**	**117**	**26,5**	**9,4**
	Difficult to categorize	**45 (4.7%)**		

Table 2 12 typical accident scenarios representing 95% of occupational injury among Swedish police officers 2005-2006 (N=964).

From Table 2 it can be seen that 95% of all accident processes could be acceptably and credibly allocated to clusters possible to differentiate in relation to exposure.

New information was generated in relation to exposures resulting in high proportions of medical impairment and long periods of sick leave. The injuries generated during training with dog and gun, and during physical training, generated the highest proportion of medical impairment. Stress-related injuries, resulting from threat to the family and needle stick, generated the longest periods of sick leave.

Prevention measures associated with job techniques should differentiate between arrests of violent suspects, mentally deranged suspects, and drunken suspects. To break glass and suffer cuts seem to be one distinct accident process, fire and the inhalation of fumes represents another type, traffic injuries and fall injuries other separate types.

In conclusion, free-text analysis shows that 95% of the known universe of police officers work-related injury risk can be described in such a way that cost-effective prevention measures can be suggested.

REFERENCES

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.

EU-OSHA (2003) **A Review and Analysis of a Selection of OSH monitoring systems** . European Agency for Safety and Health at Work, Bilbao 2003.

Field, B, Larsson, TJ, Kenningham, L & Lee, S (2000): **Falls from Heights - a report on claims, fatalities and injury severity outcomes**. Victorian Workcover Authority.

Larsson, TJ & Forsblom, M (2005) Towards an international limit value for occupational trauma risk: Industrial exposures associated with occupational trauma, permanent impairment and fatalities in a five-year national claims material. **Safety Science**, Vol 43(1) 61-71.

Hale, A., Goossens, L et al. (2004) **Managing safety barriers and controls at the workplace.** In: Spitzer, C, Shmocker, U & and Dang, V (Eds) Probabilistic Safety Assessment and Management. Berlin, Springer Verlag.

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.

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

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

Larsson, TJ (1990b) Severe hand injuries among Swedish farmers**. Journal of Occupational Accidents**, 12: 295-306.

Larsson, TJ (2003): Who Will Use Claims Data for the Prevention of Occupational Trauma? **Safety Science Monitor**, Vol 7, Issue 1.

Larsson, TJ & Lindqvist, Ch (1992) Traumatic fatalities among seafarers in Sweden 1984 - 1988**. Safety Science**, 15: 173-182.

Larsson, TJ & Field, B (2002): The Distribution of Occupational Injury Risks in the State of Victoria**. Safety Science** Vol 40(5) 419-437.

Larsson, TJ & Field, B (2002b): The Distribution of Occupational Injury Risks in the Victorian Construction Industry. **Safety Science** Vol 40(5) 439-456.

Larsson, TJ, Field, B, Sabhlok, S & McWilliams, G (2001): **The Distribution of Sprains and Strains by Gender, Age, Occupation and Industry According to the Victorian Claims Data 1996-2000.** Victorian Workcover Authority.

Larsson, TJ, Forsblom, M, Lindberg, L-G & Normark, M (2006) **Decision support for risk analysis in small enterprises.** Paper to WorkingonSafety 2006, The Netherlands.

Persson, I & Larsson, TJ (1991) Accident-related permanent disabilities of young workers in Sweden 1984-85. **Safety Science (Journal of Occupational Accidents**),14: 187-198.

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.