Evidence based policy and practice: meeting sampling challenges in evaluating national OHS interventions

Robert Pedlow
and
Jenny Job

Safe Work Australia
GPO Box 641
Canberra ACT 2601 (C220NBG)

Abstract

Purpose: Goldenhar and Schulte's 1996 review of the methodological quality of occupational intervention studies highlighted a number of problems including the selection of samples. Guidelines on impact evaluation indicate that to evaluate national intervention programs evaluators should ensure that sites and program participants are selected using probability methods to enable results to be statistically representative of population sites.
What can evaluators do when sampling frames do not exist to enable the use of probability sampling and it is not feasible to use population sampling methods to access the target workforce? Zwerling et al (1997) proposed that surveillance data be used as a means to obtain comparison groups for OHS evaluations. On that basis, in an evaluation project conducted in the road freight transport industry we investigated the feasibility of using data from a previously conducted national hazard exposure worker surveillance survey to identify a comparison group of employee drivers.

Argument: We do not dispute Kristensen's (2005:2) argument that "research quality *does* matter". However, might this mean that the research should not be conducted because a random and representative sample can not be obtained? We argue that despite difficulties with specifying sampling frames and obtaining representative samples, acceptable evaluations can be conducted and that policy makers can have confidence in the results.

Findings: We compared owner drivers and employee drivers on their demographic characteristics, manual task hazard exposures and rates of job strain. The data for owner drivers was collected as part of the evaluation study, while the data for employee drivers was taken from the National Hazard Exposure Worker Surveillance survey (Office of the Australian Safety and Compensation Council, 2008). Based on the literature we expected that these two groups of drivers would differ in several key areas. The results were broadly consistent with our expectations.

Implications for evaluation theory and practice: Government regulators and policy makers can be confident they are basing their decisions on evaluation evidence that has been validated. By using national hazard surveillance data as an additional data source we can be more confident that findings from our evaluation studies can be generalised nationally. While there is a significant and growing literature on the evaluation of occupational health and safety (OHS) interventions, very few published studies have dealt with the evaluation of national programs. One of the challenges of evaluating national programs includes obtaining representative samples of employees in the workplaces or industry sector where an intervention is taking place. In practice, sampling frames generally do not exist and due to the time and cost it is infeasible to use population sampling methods to access workers. We present an approach that we have used to address these problems; namely, using national hazard exposure worker surveillance data and data collected during the evaluation project to compare the samples. While imperfect, this approach is likely to be of interest to evaluators involved in evaluation of OHS interventions and to evaluators in other areas of government involved in the evaluation of national programs.

Introduction

As political and economic environments demand greater justification for new regulations, as well as increased accountability for existing regulations, the evaluation of government OHS policy and regulatory interventions will become increasingly important (LaMontagne, 2003). The great majority of research on the effectiveness of OHS interventions in improving workplace safety and health outcomes has focused on evaluation of workplace-level interventions. Relatively little attention has been given to evaluation of national OHS interventions. Goldenhar et al (2001), discussing OHS intervention interventions studies in the United States, acknowledge that there has been a dearth of occupational safety and health studies evaluating national policy interventions.

Along with studies of the effectiveness of OHS interventions, there is now a growing body of literature that aims to identify key methodological issues for OHS intervention research (see for example Goldenhar and Shute 1996; Zwerling, Daltroy, Fine, Johnston, Melius, and Silverstein 1997; Shannon, Robson and Guastello 1999; Goldenhar, LaMontagne, Katz, Heany and Landsbergis 2001). Reflecting the underlying focus on workplace-level intervention studies, most of the work on methodological issues has paid relatively little attention to the particular challenges involved in the evaluation of national OHS interventions. The current study addresses one of these challenges, namely the difficulties involved in obtaining representative samples for the evaluation of national OHS interventions, and presents some preliminary findings regarding a potential solution to this issue.

Before discussing the issue of representative samples for national OHS intervention research and our proposed approach in more depth, it is important to put this issue in context of the major criteria for methodological quality that have been identified in reviews of OHS intervention studies. Goldenhar and Shute (1996) identified subject selection and statistical power as one of the key methodological problems with OHS intervention studies up to that time. Kristensen (2005) argues that whether we need large representative samples depends on the particular aims of the research. Kristensen distinguishes between aetiological studies that investigate the relationship between exposures and health outcomes and prevention effectiveness studies that investigate how to effectively reduce the exposure. Kristensen argues that while for aetiological studies we require large representative samples, for prevention effectiveness studies we need well designed case studies, from a range of contexts, to test the programme theory for the intervention. We acknowledge the need to use both methods in our evaluation of OHS interventions in Australia.

Zwerling et al (1997), in discussing the evaluation of workplace interventions, note that the type of workforce chosen for the evaluation can greatly influence theextent to which the results can be generalized. Zwerling et al also recommend the implementation of hazard surveillance systems within and across companies, noting that this would greatly increase our ability to identify treatment and control groups for quasi experimental evaluation studies. Guidelines for program evaluation published by the US Government (US Government – Office of Management and Budget, nd) indicate that to

evaluate national intervention programs evaluators should ensure that sites and program participants are selected using probabilistic methods to enable results to be representative of sites and participants.

The Heads of Workplace Safety Authorities (HWSA) comprises the CEO's or their representatives of the occupational health and safety authorities within the nine Australian jurisdictions responsible for the administration for occupational health and safety laws and regulations in Australia. Each year HWSA conducts national intervention campaigns on OHS issues within particular industries. In 2007, HWSA requested that the Office of the Australian Safety and Compensation Council (now Safe Work Australia) assist with evaluation of these campaigns. The current paper presents additional analyses of data gathered in the course of work carried out by Safe Work Australia in support of the evaluation of one HWSA campaign that commenced in 2008 focusing on musculoskeletal disorders in the road freight transport industry. The paper also uses a subset of data from the National Hazard Exposure Worker Surveillance (NHEWS) survey conducted by Safe Work Australia in 2008. This study, the first of its kind conducted in Australia, aimed to gather data on hazard exposures and controls from a representative sample of workers in five priority industries: manufacturing, transport and storage, construction, agriculture, forestry and fishing, and health and community services.

The aim of all evaluations is to determine whether the intervention was effective. With national OHS interventions the aim of evaluations is to determine whether the intervention produced the desired outcome for the industry sector nationally. That is, did the intervention produce a decrease in hazard exposures and ultimately reduce deaths and injuries in the workplace? This suggests that to appropriately evaluate a national OHS intervention our sample of workers and organisations needs to be representative nationally of the industry sector. Obtaining representative samples of workers presents substantial methodological challenges. Using a population-based sample is generally infeasible due to the cost of locating sufficient respondents in one industry or occupation. As well, there are, with occasional exceptions, no sampling frames available to enable access to workers in a particular industry sectors to conduct surveys.

The present paper reports on a preliminary set of analyses that investigated the feasibility of comparing data on employee drivers from NHEWS with data on owner drivers gathered in the evaluation study. Owner drivers are for the most part not registered for workers' compensation and face substantial financial pressure that may lead them to pay less attention to OHS (Quinlan and Mayhew 1997). We hypothesised that owner drivers may be less likely to pay attention to manual task hazards in their job because they have no incentive to do so and are under pressure to complete work quickly. We also hypothesised that owner drivers would experience fewer demands as they are in charge of their own work. In summary, the aim of the analyses reported in this paper was to compare our evaluation sample of owner drivers with employee drivers from NHEWS on demographic characteristics, exposure to manual task hazards and rates of job strain.

Method

Data sources

National Hazard Exposure Worker Surveillance Survey (NHEWS): The NHEWS is the first Australian nationally representative hazard exposure worker surveillance survey conducted in Australia. It was developed and funded by the Commonwealth Government, NSW, Vic, SA, Qld, WA and NT. The survey comprised two waves. Wave one (n=1900) covered workers in five priority industries: Manufacturing; Transport & Storage; Health & Community Care Services; Construction; and Agriculture, Forestry & Fishing. Wave two (n=2600) comprised a mix of priority and non priority industries. Data collection for the NHEWS survey was conducted using Computer Assisted Telephone Interviewing (CATI) and carried out by an independent market research company, Sweeney Research Pty Ltd. Data collection for this study was carried out between January and July 2008. Ethics review and approval for the research was obtained through the University of Sydney's Human Research Ethics Committee. A 42% response rate for the survey was obtained. To be eligible to take part in the study participants had to have undertaken paid work in the last week. The study collected data from workers on their self reported exposures to a wide range of workplace hazards including noise, vibration, manual tasks, psychosocial, sun, chemical, airborne and biological hazards. In addition, the study collected data on use of controls for these hazards, as well as basic demographic information.

Road Freight Transport Evaluation: As part of its role in support of the evaluation of the 2008-09 HWSA Campaign on manual tasks in the road freight transport industry, Safe Work Australia commissioned Sweeney Research to administer surveys to owners or managers of road freight transport businesses, and owner drivers and employee drivers in the road freight transport industry in NSW, QLD, WA, TAS and SA. The research was conducted between 4 December 2008 and 22 January 2009. A total of 151 owners or managers (owners/managers), 198 owner drivers, and 154 employee drivers in the road freight transport industry completed surveys. The initial sample of owner/managers and 93 of the 208 owner drivers were obtained from the Dunn and Bradstreet list of Australian Businesses (Dunn and Bradstreet, nd). We sampled randomly from the subset of business in this directory that were listed as being in ANZSIC Code 6110: road freight transport. This group was surveyed using CATI. The employee drivers and the balance of the owner driver group were obtained from an online research panel and were surveyed online. The study collected data on:

- hazard exposures including:
    - manual task hazards
    - hazards specific to the road freight industry – eg traffic hazards during loading and unloading of vehicles
    - job demand and control (Karasek and Theorell 1990)
- safety climate (Loughborough University, nd, p.46)
- use of controls
- drivers' health behaviour

- demographic information.

Due to the low response rate for the online survey, for this paper we only use data for the 93 owner drivers who were surveyed using CATI. There was a 20% response rate for the CATI survey overall.

Data Analysis

Comparison method: In previous work undertaken as part of the NHEWS survey, verbatim responses given by participants regarding their occupation were coded for industry and occupation using the ANZSCO system. As a result, for the present study we were able to identify 89 truck drivers in the transportation and storage industry from the NHEWS database. We compared this group with 93 owner drivers in the evaluation study who were surveyed using CATI. To compare the two groups we constructed a new file with the data for the 93 owner drivers from the evaluation study and the 89 drivers from NHEWS. This file contained the variables from both studies that were identified as being precisely equivalent in the two surveys. With one exception, the wording of the question and the response scale for the items were the same in both studies. Because the response scale for the consequence of physical demands differed between the two studies we dichotomised the scale with 0 being 'did not report any consequence of the physical demand' and 1 being 'reported consequence of the demand'. Rather than directly reporting the rates of reported consequences of physical demands we compared the associations between exposure to manual tasks and experiencing consequences of physical demands in the two datasets.

Job strain measure: The responses to the demand and control scales were analysed using the method given by LaMontagne, Keegel, Vallance, Ostry and Wolfe (2008).

Results

We compared the two samples on demographic variables, exposure to manual task hazards, effects of exposure to manual task hazards on physical symptoms, and rates of job strain. There were no significant differences between the two samples on age or gender. Drivers were predominantly male, with 93% males in the NHEWS group and 96% males in the evaluation study. The majority of drivers in both the evaluation sample and NHEWS were aged 35 years and over.

The evaluation sample had a significantly lower level of education than the NHEWS sample. 82% of the evaluation group had Year 12 or lower levels of education while the comparable figure for NHEWS was 37%. The NHEWS sample comprised mostly employee drivers (85%) with 15% having their own business or contracting. The evaluation sample comprised entirely owner drivers who were solo operators.

Table 1 – Demographic characteristics of the NHEWS and evaluation sample

	NHEWS (n=89)	Evaluation sample (n=93)	X^2	*p*
Demographic variables				
Age				
15/18-24	2	-		
25-34	7	2		
35-44	29	25		
45-54	37	36		
55+	25	37		
			7.0	ns
Gender				
Male	93	96		
Female	7	4		
			0.3	ns
Level of Education				
Year 12 not complete	35	57		
Year 12 complete	2	25		
Trade cert/TAFE	56	10		
Uni degree	1	7		
Other	5	1		
			62.3	p<.001
Employment arrangement				
For an employer	85	na		
Own business	7	na		
Contractor	7	na		
Owner driver – own business – solo operator	1	93	-	-

It is possible that the difference in educational level between the two samples is at least in part a consequence of one sample comprising owner drivers while the other comprised mostly employee drivers. This suggests that truck companies may be increasingly either employing drivers with trade certificates and/or supporting their drivers to complete trade certificates.

Manual task hazard exposure: The drivers in the NHEWS sample were significantly more likely to report exposure to all of the manual task hazards compared to the drivers in the evaluation sample. For example, 68% of drivers in the NHEWS sample reported being exposed to carrying or lifting heavy objects compared to 46% of drivers in the evaluation study. 75% of drivers in the NHEWS sample reported being exposed to work with their body bent forwards compared to 30% in the evaluation study.

Table 2 – Manual task hazard exposures in NHEWS and the evaluation sample

	NHEWS (n=89)		Evaluation sample (n=93)		X^2	*p*
Manual task hazard exposures	Not Exp.	Exposed	Not Exp.	Exposed		
Carry or lift heavy Loads	32	68	54	46	9.3	p<.001
Make the same hand/arm movements over and over again	2.2	98	39	61	36.6	p<.001
Work with your body bent forwards	25	75	70	30	37	p<.001
Twisted or awkward posture	36	64	59	41	9.8	p<.005
Work with your hands raised above our head	30	70	55	45	11.1	p<.005
Work while sitting down	3	97	28	72	20.5	p<.001
Squatting or kneeling while you work	31	69	60	40	15.1	p<.001
Push or pull using some force	17	87	57	43	31	p<.001
Work standing in one place	49	50	43	57	0.8	ns

There was a time difference of approximately 6 months between the collection of the NHEWS sample and the evaluation sample. It is possible that increased use of controls in the industry or other changes might explain the differences seen. For example, a downturn in business could have reduced drivers' exposure to manual task hazards during this time. We believe it is more likely that the difference is due to differences between the owner drivers surveyed in the evaluation study and employee drivers who make up the majority of the NHEWS sample. The finding supports our hypothesis that owner drivers may be less likely to report hazard exposures. This may be for two reasons: the added financial pressure many are under, and the fact that they are as a group usually self employed and may pay less attention to OHS.

We analysed the association between high levels of exposure to manual task hazards and experiencing the consequences of physical demands separately for the two samples. We defined high levels of exposure to manual task hazards as being exposed to more than four manual tasks. There is a consistent pattern in both groups of high levels of exposure to manual task hazards being associated with experiencing the consequences of physical

demands. However, the association between manual task hazard exposure and pain in the hips, legs, knees and feet did not reach statistical significance for the NHEWS sample.

Table 3 – High manual task hazard exposures by pain and tiredness

	NHEWS (n=89)				Evaluation sample (n=93)			
Manual task hazard exposure	Low	High	X^2	*p*	Low	High	X^2	*p*
Experience consequence of physical demands								
Tiredness								
Never	20	0			58	11		
Rarely-all the time	80	100			42	89		
			5.62	p<.05			21.1	p<.001
Pain – back or neck								
Never	39	4			65	36		
Rarely-all the time	61	96			35	64		
			9.96	p<.005			7.35	p<.01
Pain - shoulders/ arms wrists hands								
Never	49	8			60	36		
Rarely-all the time	51	92			40	64		
			12.4	p<.001			4.89	p<.05
Pain – hips, legs, knees or feet								
Never	45	29			61	36		
Rarely-all the time	55	71			39	64		
			1.74	ns			5.65	p<.05

The results in Table 3 suggest that there is a consistent pattern of association between manual task exposure and experiencing the consequences of physical demands across the two samples. This supports the external validity of the evaluation as it suggests that the same process is operating in the two samples.

The analyses shown in Table 4 below indicate that drivers in both the evaluation and NHEWS samples were not experiencing job strain. Job strain is defined as the 25% of workers who experience high demand and low control over their work. Nevertheless, more drivers in the NHEWS sample were in the low strain quadrant while more drivers in the evaluation sample were in the passive quadrant. Drivers in both samples experienced similar levels of control however, drivers in the evaluation sample were more likely to experience low demand.

It is likely that the difference between the two samples is due to the NHEWS sample comprising mostly employee drivers and the evaluation sample comprising owner drivers. The items in the demand scale are concerned with time pressure, unachievable demands and having to multi-task. It is likely that owner drivers may experience their job as less demanding because of they are in charge of their own work.

Table 4 – Psychosocial hazard exposures in NHEWS and the evaluation samples

	NHEWS (n=89)	Evaluation sample (n=93)	X^2	*p*
Job Quadrant				
Passive	22.5	54.8		
High Strain	6.7	18.3		
Low strain	46.1	14.0		
Active	24.7	12.9	36.18	p<.001
High demand	31	0		
low demand	69	100	33.0	p<.001
High control	73	71		
low control	27	29	0.7	ns

Discussion

The aim of this paper was to investigate the feasibility of comparing data on employee drivers from a previously collected National Hazard Exposure Worker Surveillance survey with data on owner drivers gathered in the evaluation study. We compared the demographic characteristics and manual task and psychosocial hazard exposures from a group of owner drivers surveyed as part of the evaluation of an intervention campaign aimed at reducing musculoskeletal injuries in the road freight transport industry with a group comprised primarily of employee drivers selected from the NHEWS survey data set. The two groups had similar age and gender profiles, with the owner drivers having lower levels of education. Owner drivers reported lower levels of exposure to manual task hazards than employee drivers. The relationship between exposure to manual task hazards and experiencing the consequences of exposure to manual tasks demands was similar in the two samples. Owner drivers experienced lower levels of demand compared to employee drivers.

It is important to acknowledge the potential limitations of comparing surveillance data with data gathered for an evaluation project. If there is a time difference between the conduct of the evaluation study and the collection of the surveillance data, it is possible that other changes may have occurred in the industry sector during that time, for example, high profile accidents, or other activity by regulatory authorities. The time of year may have an impact in some industry sectors eg. the work profile may be different in the lead up to Christmas compared to mid year. Also, measures may not be precisely parallel between the evaluation study and the surveillance data. For example, response scales may not be the same or item wording may have changed. This reinforces the desirability of using standardised measures where possible. However, to the extent that we can identify consistent relationships between constructs across studies this can support convergent validity even if the measures are different. Finally, large scale surveillance data will necessarily be more restricted than that needed for evaluation purposes. We suggest that the ideal approach may be for evaluation studies to collect more

detailed information from whatever groups of workers can be obtained while using surveillance data to explore specific quantitative hypotheses.

Following Kristensen (2005), we argue that the methodological requirements for evaluation of interventions depend to some extent on the nature of the evaluation questions we are seeking to answer. To the extent that we are seeking to determine whether an intervention is resulting in improved health and safety outcomes, where these are defined as reduced rates of deaths, injuries and disease, then large representative samples and other rigorous methodological controls are required. To the extent that we are seeking to determine whether an intervention will produce a change in behaviour in the workplace, we need case studies from a range of different contexts that enable us to develop and test program theory for the intervention. Green and Tones (1999), in discussing the evaluation of public health interventions, argued that rigorous experimental and quasi-experimental designs, while important, are not able to answer all evaluation questions and that qualitative data grounded in program theory should also be considered.

In relation to the evaluation of national OHS interventions, policy makers will require the answer to both kinds of evaluation questions. That is, do the interventions reduce hazard exposures in the workplace, and also, do they result in reduced rates of deaths and injuries in the workplace? The findings in this paper suggest that national hazard exposure surveillance data can be used to test specific quantitative hypotheses for evaluation studies. We see this as a potentially useful approach to evaluation of national OHS interventions while acknowledging that we will also require richer data sources for evaluations. These data will have to be collected using available samples. We plan to further explore identification of specific groups of workers from surveillance data and to use case studies to gather qualitative data based on program theory. We consider that doing so can enable us to better understand how interventions work to reduce hazard exposures.

References

Dunn and Bradstreet, (nd). Dunn and Bradstreet list of Australian Businesses. www.dnb.com. Accessed 4/8/2008.

Goldenhar, L.M., LaMontagne, A. D., Katz, T., Heaney, C., Landsbergis, P. (2001). The intervention research process in occupational safety and health: An overview from the National Occupational Research Agenda intervention research effectiveness team. *Journal of Occupational and Environmental Medicine*, 43, 616-622.

Goldenhar, L. M and Shute, P. A. (1996). Methodological issues for intervention research in occupational health and safety. *American Journal of Industrial Medicine*, 29, 289-294.

Green, J. and Tones, K. (1999). Towards a secure evidence base for health promotion, *Journal of Public Health Medicine*, 21, 133-139.

Karasek, R., and T. Theorell, T. (1990). *Healthy work*. Basic Books, New York.

Kristensen, T. S. (2005).Intervention studies in occupational epidemiology. *Occupational and Environmental Medicine*, 62, 205-210.

LaMontagne, A. D. (2003). Improving occupational health and safety policy through intervention research. National Research Centre for OHS Regulation Working Paper 18.

LaMontagne, A. D., Keegel, T., Vallance, D., Ostry, A., and Wolfe, R. (2008). Job strain – attributable depression in a sample of working Australians: Assessing the contribution to health inequalities. *BMC Public health*, 8, www.biomedcentral.com/1471-2458/8/181.

Loughborough University. (nd). Safety Climate Measurement User Guide and Toolkit. http://www.lboro.ac.uk/departments/bs/safety/document.pdf. Accessed 4/8/2009.

Mayhew, C., Quinlan, M. and Ferris, R. (1997). The effects of subcontracting/ outsourcing on occupational health and safety: Survey evidence from four Australian industries. *Safety Science*, 25, 163-178.

Office of the Australian Safety and Compensation Council (2008). National Hazard Exposure Worker Surveillance Survey 2008 http://www.safeworkaustralia.gov.au/swa/aboutus/publications/2008researchreports.htm . Accessed 5/8/2009.

Shannon, H.S., Robson, L.S., Guastello, S. J. (1999). Methodological criteria for evaluating occupational safety intervention research, *Safety Science*, 31, 161-179.

US Government - Office of Management and Budget, (nd) Assessing program performance. www.whitehouse.gov/OMB/part. Accessed 3/8/2009.

Zwerling, C. Daltroy, L.H., Fine, L. J., Johnston, J.J., Melius, J., Silverstein, B. (1997). Design and conduct of occupational injury intervention studies: A review of evaluation strategies, *American Journal of Industrial Medicine*, 32, 164-179.