

Measurement Properties of Occupational Health and Safety Management Audits: A Systematic Literature Search and Traditional Literature Synthesis

Lynda S. Robson, PhD,¹ Philip L. Bigelow, PhD^{1,2}

ABSTRACT

Objective: The measurement properties of Occupational Health and Safety (OHS) management audits might be important in some applications, especially when audit scores are treated as performance measures. The review, therefore, sought to identify and summarize the research evidence on the measurement properties (e.g., reliability, validity) of methods of OHS management audit.

Methods: Bibliographic databases in business, medicine and OHS were systematically searched. Evidence from relevant publications was synthesized using traditional narrative review methods.

Synthesis: The literature on the measurement properties of OHS management audit methods is sparse. Seventeen relevant audit methods were identified. Content validity was demonstrated for only five audit methods. Inter-rater reliability was formally tested for only three audit methods and construct validity for only one. There were no studies of test-retest reliability or responsiveness. The investigations of inter-rater reliability (i.e., consistency among auditors) showed that it is often unacceptably low.

Conclusion: There is a research gap concerning the measurement properties of OHS management audit methods. The available research raises questions about the properties of audit methods in current use.

Key words: Accident prevention; management audit; occupational health; reproducibility of results; safety management; validation studies

La traduction du résumé se trouve à la fin de l'article.

Can J Public Health 2010;101(Suppl.1):S34-S40.

An occupational health and safety (OHS) management system is designed to protect the health of workers by the following means: designating roles and responsibilities related to OHS; setting organizational targets and objectives related to OHS; planning and establishing the maintenance of hazard controls; and monitoring, reviewing and improving the system's implementation and effectiveness. Federal and provincial legislation (e.g., the *Occupational Health and Safety Act* in Ontario) specifies simple management systems applicable to all workplaces. Exemplary organizations also seek compliance with voluntary standards and guidelines¹⁻³ that are more comprehensive.

Auditing is a means of directly and comprehensively monitoring the implementation and effectiveness of a firm's OHS management system.⁴ The auditing process typically involves the following:⁵ gathering evidence about the management system through interviews, documentation reviews and work site observations, guided by an audit instrument; evaluating the gathered evidence; and providing a summary of the evaluative findings. Auditing is an important component of an OHS management system,¹⁻³ and up to 95% of Fortune 2000 companies perform audits.⁶ While less prevalent in smaller firms, audits are recognized as nevertheless relevant.⁷

Depending on the type of application in which a particular OHS management audit is involved, researchers and practitioners might want to consider its measurement properties (i.e., reliability, validity, etc.). In some cases, measurement properties are relatively less important. For example, simple audits are sometimes used to assess

the management and programmatic needs of organizations that are just starting to develop their OHS management systems.⁸ Similarly, there are firms in which management systems are more developed but in which audits are used periodically only to ensure that there are no major gaps in the management system. In these two situations, only a blunt measurement instrument is needed, though one would want some assurance of its content validity: one would want to be confident that the key elements of an OHS management system, appropriate for the organization, are adequately represented in the content of the audit instrument, otherwise it cannot serve its function of identifying gaps. In contrast to the previous examples, additional measurement properties (e.g., inter-auditor reliability, predictive validity, responsiveness) might also be quite important when audits are used in performance measurement applications. These include benchmarking, determining whether a particular standard has been met or monitoring progress over time. Measurement properties may be additionally important in such applications when the results of audits determine organi-

Author Affiliations

1. Institute for Work & Health, Toronto, ON

2. University of Waterloo, Waterloo, ON

Correspondence and reprint requests: Lynda Robson, PhD, Institute for Work & Health, 481 University Ave., Suite 800, Toronto, ON M5G 2E9, Tel: 416-927-2027, ext. 2164, Fax: 416-927-4167, E-mail: lrobson@iwh.on.ca

Acknowledgement: The Institute for Work & Health receives core funding from the Workplace Safety & Insurance Board of Ontario (WSIB). The review was also supported in part by the WSIB Prevention Reviews initiative. The authors appreciate the key support of the library staff, particularly Emma Irvin and Rachel Couban, and the administrative assistance of Diana Pugliese.

Conflict of Interest: None to declare.

zational rewards, e.g., financial,⁹⁻¹² recognition,¹³ contract opportunities^{10,14} or even penalties.¹⁵ To the extent that the data provided by the audit of a firm are not reliable, valid or responsive, organizational actions based on the audit report might be misdirected.

Our original intent was to review the research evidence on the measurement properties of OHS management audits using the methods of systematic literature reviews. However, initial screening revealed a lack of published studies, and we therefore only used systematic review methods for searching and screening the literature, and then used traditional review methods to assess and synthesize the evidence.

METHODS

Literature search

The search strategy was developed in MEDLINE, starting with the Medical Subject Heading (MeSH) terms assigned to known, relevant articles. This search strategy was applied to five other bibliographic databases, adaptations being made as required: American Business Inform (ABI), CCInfoWeb, Econlit, EMBASE and Health and Safety Science Abstracts. ABI and CCInfoWeb catalogue grey literature sources, as well as those that are peer-reviewed. The databases were searched from their point of inception until September 2008. One search strategy looked for abstracts classified with the MeSH term "Management Audit" AND one of the following terms: "Wounds and Injuries", "Accidents, Occupational", "Accident Prevention" or "Occupational Health". The second, more fruitful, strategy looked for abstracts classified with the MeSH term "Safety Management" AND the free text term "Audit." No restrictions were placed regarding date and language of the original publication. In order to broaden the search, the following sources were also used: reference sections of publications deemed relevant, personal files of the authors, and an in-house bibliographic database assembled for a systematic review of OHS management system effectiveness.¹⁶

Relevance screening

The titles and abstracts arising from the searches were reviewed to identify potentially relevant publications using two inclusion criteria:

- The publication is a journal article, book, conference proceeding, dissertation or report,
- The publication contains information on any of the following measurement properties of OHS management audits: content validity, construct validity, inter-rater reliability, test-retest reliability or responsiveness.

Nine references¹⁷⁻²⁵ were eliminated when the following exclusions were applied:

- Publication is a magazine article or newsletter
- Audit focuses on hazards rather than on management^{17,18}
- Audit focuses on the management of a particular type of OHS hazard instead of all OHS hazards¹⁹
- Audit is a safety management system audit, which does not focus on OHS²⁰⁻²⁵

The last exclusion pertains to a distinct stream of research and practice focused on the prevention of out-of-control processes or catastrophic events. Some of this literature was reviewed in an earlier report,²⁶ and the basic conclusion drawn from it is similar to the one we draw for this review.

The review of the titles and abstracts was shared by the authors, each title and abstract being reviewed by a single author. Potentially relevant publications were retrieved and reviewed in more detail.

Evidence extraction and synthesis

Both authors were involved in the data extraction and evidence synthesis. Each author read all retrieved publications, discussed and developed a common understanding of the findings, and agreed upon which publications did not meet the criteria; they shared the extraction and synthesis of the evidence.

RESULTS

Table 1 summarizes the features of 17 distinct audit methods found in the relevant publications. Several well-known audit methods were identified through the review: the Diekemper and Spartz method,²⁷⁻²⁹ the International Safety Rating System (ISRS)³⁰⁻³⁵ and the CHASE audits.^{33,36} Another two were associated with the American Industrial Hygiene Association.^{37,38,41-45} Several of the audits were intended for multiple sectors of the economy, whereas others target a single sector. The number of items in each audit instrument is in the range of less than 100 to several hundred. Some methods require the auditor to respond to an item by indicating yes or no; other methods allow a greater variety of responses. Less commonly, auditors are asked to assign a number of points out of a maximum possible number. The most common way of summarizing the overall audit results is a percent score (out of 100).

For eight of the methods listed in Table 1, either we considered the content validity to have been evidenced, or a formal test of reliability or validity had been reported. This evidence is summarized in Table 2 and is discussed further below. For the remaining nine audit methods, only information of a more preliminary nature was available, and they are not discussed further.

Content validity

In the present context, content validity⁵⁶ refers to the comprehensiveness of the audit instrument in its representation of management system concepts. There are five methods for which content validity is demonstrated. The first^{37,38} had ISO 9001 as the organizing framework and drew its content from several OHS and environmental management system documents. A second method,⁴¹⁻⁴⁵ also developed at the University of Michigan, defined a "universe" of OHS management system elements. Researchers reviewed 13 OHS and environmental standards or guidance documents and selected four that collectively represented the content of all 13. These four "input models" were deconstructed and then reorganized into an integrative model. A third method was developed in collaboration with the Canadian Pulp and Paper Association⁴⁷ using a matrix approach to the audit content: 11 OHS activity areas (e.g., emergency responsiveness preparation, health surveillance) were each considered in relation to the presence of key management system components (goals and procedures, assignment of responsibility, performance measurement, review of performance, corrective action). A fourth method, developed in Norway for the mining sector,^{39,40} was intended for consensus-based, high-level, self-audits by organizational representatives. The final example comes from a Singapore project in the construction sector,⁵⁴ in which the steps involved included reviewing existing methods, generating items,

Table 1. Audit Methods Included in the Review*

Name of Audit Method*	Target Population	Conceptual Basis	Number of Items	Nature of Response Options	Form of Final Output
Diekemper & Spartz method ^{27,29}	Manufacturing	Unknown	29	1 to 4 (poor to excellent)	Percent (of total possible points)
International Safety Rating System (ISRS) – Mining ^{30,32}	Mining operations in South Africa	Unknown. Mining version adapted from generic version by adding material. ³⁰	“almost 1000” ³¹	Mainly yes/no; some 0-20 scales	Numerical score and star rating. Star rating has five standard levels (no. of stars 1-5) and five advanced (no. of stars 1-5). Each level involves a different number of elements. See ISRS – Mining
ISRS – Generic, 4th ed. ^{3,2,35}	General industry, with first version developed in the steel industry (Eisner & Lege ³⁰)	Unknown	From about 100 to 627, depending on the star level attempted ³³	Mainly yes/no; some 0-20 scales	Percent (of total possible points) overall and by section
CHASE series ^{3,3,36}	CHASE-I and -II: general industry (<100 and 100+ employees) Construction-CHASE: all sizes COSH-CHASE: any size general industry with chemical exposures	Based on review of regulations (UK) and other sources, professional judgment and input from pilot testing. Weights assigned according to predicted risk if hazard were not controlled.	CHASE-I: 200 CHASE-II: 400 Construction-CHASE: 350 COSH-CHASE: 420	Yes/no (but software can also accept a graded response such as %)	
AIHA ISO 9001 harmonized ^{37,38}	All sectors	ISO 9001	135 clauses (determined by review authors based on Dyjack ³⁷)	In conformance, minor non-conformance	Qualitative: Non-conformance reports that identify deviation from standard and its root cause Percent (of total possible points). Also, list of significant items requiring action.
Goodyear Tire and Rubber Co. – Industrial Health Review and Corporate Safety Systems Review; 1995 versions ^{37,38}	Manufacturing	Unknown	At least 191 (based on the combined number of subelements in the two reviews)	Each subelement worth a number of points. Auditor awards full or partial points.	
Safety Element Method ^{39,40}	Norwegian mining	Principles of safety management and quality assurance. Action research process used to develop content, with experts referring to ISRS and other existing tools.	6 elements + 12 subelements	Each subelement and element scored from stage 1 to stage 5 (stage 1 = no formal system; stage 2 = minimum regulated standards; stage 5 = fully developed)	Categorical assessments for each subelement and element. Also, can convert results to score by summing items, with item score corresponding to the stage ⁴⁰
AIHA Universal OHSMS Assessment Instrument ⁴¹⁻⁴⁵	All sectors	Reconstructed from four input models, using system theory and policy analysis models for organizing framework.	486	Ordinal score for each measurement criterion (0-5) in most recent version ⁴⁵	Average score (0-5)
Safety Management Audit for Construction (SMA-CON) ⁴⁶	Construction in Hong Kong	Based on British Standard 8800:1996 and elements specified by Labour Department.	456 (less for firms with <10 employees)	Yes/no and assigning points out of 20 or 30.	Percent (of total possible points)
Canadian Pulp and Paper Association ⁴⁷	Canadian pulp and paper	General OHS practices/principles and management system principles. 11 OHS activities assessed with respect to 5 management system elements.	55	Score 0 or 1	Scores summed by activity or by management system component
Water utilities benchmarking audit ⁴⁸	All sectors	Principles from British Standards Institute and the Health and Safety Executive	56	Scores from 1 (low) to 4 (high)	Percent (of total possible points) overall and by section
Method for Industrial Safety and Health Activity Assessment (MISHA) ²⁸	Manufacturing	Framework based on Booth and Lee ⁴⁹ key elements of safety management	55	Scores from 0 (not acceptable) to 3 (no weakness; strong improvement process in place)	Percent (of total possible points) overall and by section
AS/NZS 4804-based ^{50,51}	Small and medium enterprises in metal prefabrication	Based on Australian/New Zealand standard, AS/NZS 4804	80	Yes=1; no=0; partial=0.5	Percent (of total possible points) overall and by section

...continues/

Table 1. continued...

Name of Audit Method*	Target Population	Conceptual Basis	Number of Items	Nature of Response Options	Form of Final Output
International Truck and Engine Corporation ⁵²	Single, multi-site truck and engine manufacturer	Adapted from Goodyear Tire and Rubber Company audit; expanded to include health promotion & wellness.	64 elements (there are 64 elements and there may be subelements too)	Each element/subelement is numerically valued, and auditor awards full or partial numerical credit for each.	Percent (of total possible points) overall and by section Also, list of non-compliances and uncontrolled hazards. Score out of 100
Wellworks-2 audit ⁵³	US manufacturing	Based on the US OSHA Program Evaluation Profile categories	91	Yes/no	Overall mean scores out of 1.0 for each factor and for total
Construction Safety Index ⁵⁴	Construction industry in Singapore	Factors derived from factor analysis: policy, process, personnel and incentive	590 attributes	Some attributes assessed as 0 (not in compliance) or 1 (in compliance). Others assessed on scale ranging from 0 to 1, where the value is the ratio of # of sampling units in compliance to the total number of sampling units selected.	
OHSAS 18001-based ⁵⁵	Manufacturing	Based on OHSAS 18001	43	Categorized as OK, not OK, or not documented.	Percent of "not OK" and "not documented"

* Audit methods are listed in chronological order by date of the first publication cited.

Table 2. Summary of the Research Evidence on the Measurement Properties of OHS Management Methods

Audit Method	Content Validity*	Inter-auditor Reliability† Test 1: – Test 2: ±	Construct Validity‡
D&S method ²⁷⁻²⁹			
ISRS – Mining ³⁰⁻³²			–
AIHA ISO 9001 harmonized ^{37,38}	+		
Safety Element Method ^{39,40}	+		
AIHA Universal OHSMS Assessment Instrument ⁴¹⁻⁴⁵	+	–	
Canadian Pulp and Paper Association ⁴⁷	+		
MISHA ²⁸		Original: – Revised: +	
Construction Safety Index ⁵⁴	+		

* Classification of the evidence on content validity: +, content validity is considered adequate by the review authors because the method of developing the content for the audit instrument is well described and has involved appropriate stakeholders; –, a formal test of content validity in the research literature concludes that the audit method is not valid. Cell is left blank when the available information is not sufficient for reviewers to assign either + or –.

† Classification of the evidence on inter-rater reliability: +, test(s) of inter-rater reliability yield a Kappa statistic that is 0.41 or more (categorical data) or an intraclass coefficient that is 0.75 or more (continuous data); –, test(s) of inter-rater reliability yield a Kappa statistic that is less than 0.41 or an intraclass coefficient that is less than 0.75. Cell is left blank when no test has been conducted.

‡ Classification of the evidence on construct validity: +, the majority of hypotheses constructed about the expected relations between audit data and other OHS measure(s) in a test of construct validity are supported; –, the majority of hypotheses constructed about the expected relations between audit data and other OHS measure(s) in a test of construct validity are not supported. Cell is left blank when no tests of construct validity have been conducted.

surveying employers regarding item importance, identifying organizing categories through factor analysis, and determining scoring weights through systematic consultation with experts.

Inter-rater reliability

Inter-rater reliability is the consistency of assessments of the same workplace(s) by different auditors. For categorical data, agreement is best expressed using kappa or weighted kappa (κ_w),⁵⁷ which should minimally be 0.41.^{58,59} For continuous data, intra-class coefficient statistics are used, with values of 0.75 considered minimal.⁵⁹ Inter-rater reliability was investigated for three audit methods as part of thesis dissertations.

Kuusisto²⁸ studied the reliability of a version of the Diekemper and Spartz method.²⁷ He found poor reliability when his own ratings were compared with those of local company evaluators (κ_w s of -0.03 to 0.46). Agreement between raters was better when Kuusisto's ratings were compared with those of his safety specialist students (κ_w s of 0.36 to 0.83). Kuusisto²⁸ also examined the reliability of his own Method for Industrial Safety and Health Activity Assessment by comparing his ratings with those made by members of the company being audited. Weighted kappas were all less than 0.41, so the tool was revised. A test of the new version, which had modified guidance for decision making, showed improvement (κ_w s of 0.38 to 0.58).

Dyjack et al.⁴¹ looked at the reliability of a portion of the Michigan Universal Assessment Instrument. The agreement between raters was considered inadequate by several statistical criteria, in spite of the expert qualifications of the two auditors in the project, the similarity of their substantial preparation to use the audit and the content validity of the instrument.

Construct validity

The construct validity of an audit method is established by making hypotheses about the expected relations between audit scores and other measures of OHS, testing the relations using appropriate methods and then confirming the majority of the expectations.^{56,57} There is only one study of construct validity in the literature on OHS audits that uses statistical analyses. Eisner and Leger³⁰ examined the correlation of the number of stars awarded by the ISRS audit with each of fatality rate and reportable injury rate. Correlations were small, not all in the

expected direction, and none were statistically significant. However, the study had significant methodological limitations.

Other measurement properties

There were no formal investigations found of test-retest reliability (consistency of results upon repeat administration of the audit) or responsiveness (ability to accurately detect change over time).^{56,57}

DISCUSSION

Principal findings

The research literature on the measurement properties of OHS management audits is sparse. There has been some investigation of inter-rater reliability, which showed that it was often unacceptably low according to usual standards.

Strengths and limitations of the review

This is the first literature review about the measurement properties of OHS management audits. An extensive, systematic search of the literature was conducted. We are confident that our portrayal of the research literature as sparse in this area is valid. We note, however, that there may be relevant information, particularly on proprietary methods, in sources outside of the research literature. Although this additional information would expand upon what is known from the research literature, it is not feasible to collect it in a systematic way.

Review findings in relation to other research

There are no other reviews of this nature in the research literature, preventing comparison with a similar study. However, others have commented on the paucity of available research. In 1988, Eisner and Leger^{30,p.143} remarked that, "A thorough search ... failed to discover any publication evaluating the [ISRS] scheme" by academic authorities, even though the audit method was internationally recognized at the time. One decade later, Dyjack^{37,p.80} said he was "unable to identify published studies evaluating the accuracy and repeatability of either publicly or privately held occupational health and safety assessment instruments."

There is an additional study of inter-auditor reliability known to the authors but not included in the review because it has only been reported as a conference abstract.⁶⁰ It found acceptable levels of agreement between two qualified auditors for only one of three organizational units audited.

Practical implications of review findings

On the basis of our findings, we expect that some audit instruments in common use have low inter-rater reliability. This could be a concern when various auditors conduct audits within an auditing program and the audit scores are used as a performance measure. Low inter-rater reliability under these circumstances might lead, for example, to inconsistent classification of whether firms meet the performance benchmark, erroneous ranking of firms in intra- and inter-firm comparisons, and erroneous trend information. Our findings indicate a lack of information on the construct validity and responsiveness of audit scores. Researchers, therefore, still need to establish the extent to which decision makers should rely on audit scores as an indicator of firm OHS performance. Our recent study of audit methods⁸ indicates that practitioners use them for performance measurement purposes, even in the absence of formal

knowledge about their measurement properties. We encourage a practitioner who relies on an audit method for such purposes to seek more information about its measurement properties and in the absence of that information be cautious in their decisions based on audit results.

Future research

This review points to a large research gap regarding the measurement properties of OHS management audits. There is also little empirical information on the extent to which various factors affect these properties, since there are only single case reports^{28,61} and prescriptive information from experts.⁶²⁻⁶⁷ According to these and the research literature on financial audits,⁶⁸⁻⁷⁷ we would expect determining factors to fall into the following categories: 1) auditors (e.g., their independence, their expertise), 2) auditing programs (e.g., training processes, quality control processes) and 3) the auditing method (e.g., content, usability of the audit instrument, number of response categories in the audit instrument, explicitness of decision aids for auditors). Further, it would be interesting to know, through an intervention study in the field, the degree to which measurement properties can be improved. On the other hand, the degree to which such change would affect decision making and at what cost also remains to be determined.

There are many practical challenges involved in research on management audits: they are labour intensive for the auditing organization and the workplace, the methods involved in studying audits (e.g., presence of a researcher or a second auditor) could have an impact on the audit results that is difficult to measure, sample size may be constrained by the volume of an audit program, and other data sources available for construct validity studies (e.g., injury reports) may be of poor quality. However, the studies involved in this review provide some examples of what is possible. OHS researchers could also look to the clinimetrics field for further guidance in and illustrations of high-quality measurement research.^{56,57,59,78-80}

REFERENCES

1. Canadian Standards Association. CSA Z1000-06. Occupational Health and Safety Management. Mississauga, ON: CSA, 2006.
2. International Labour Organization. Meeting of Experts on ILO Guidelines on Occupational Safety and Health Management Systems: Final Report. Geneva: International Labour Organization, 2001.
3. OHSAS 18001:2007. Occupational Health and Safety Management Systems – Requirements. London: OHSAS Project Group, 2007.
4. Karapetrovic S, Willborn W. Generic audit of management systems: Fundamentals. *Managerial Auditing J* 2000;15(6):279-94.
5. Waring A. *Safety Management Systems*. London: Chapman & Hall, 1996.
6. Nash JL. Who is auditing your safety auditors? *Occup Haz* 2005;67(7):31-34.
7. Grant J, Brown D. The inspector cometh. *Cdn HR Reporter* 2005;18(2):13,17.
8. Robson LS, Macdonald S, Van Eerd D, Gray G, Bigelow P. Prevention System OHS Management Audit Methods: Description, Content Validation and an Assessment of the Feasibility of Measurement Research. Final Report to WSIB RAC on Project #06112. Toronto, ON: Institute for Work & Health, 2008.
9. Workplace Safety & Insurance Board of Ontario. The Safety Groups Program. Available at: <http://www.wsib.on.ca/wsib/wsibsite.nsf/public/SafetyGroupsProgram> (Accessed February 13, 2009).
10. Government of Alberta. What is a Certificate of Recognition? Available at: <http://employment.alberta.ca/cps/rde/xchg/hre/hs.xml/334.html> (Accessed February 13, 2009).
11. WorkSafe BC (the Workers' Compensation Board of BC). Partners Program (COR). Available at: http://www.worksafebc.com/insurance/partners_program/default.asp (Accessed February 13, 2009).
12. Workplace Health, Safety & Compensation Commission of Newfoundland and Labrador. PRIME. Frequently asked questions. Available at: http://www.whscc.nl.ca/prime/PR_FrequentlyAskedQuestions.whscc (Accessed February 13, 2009).

13. EUSA. Zeroquest. Introduction. Available at: http://63.135.106.115/Home.aspx?PageID=13&mid=_ctl0_MainMenu2__ctl1-menuItem000 (Accessed February 13, 2009).
14. NEPCON. OHSAS 18001 certification. Available at: <http://www.nepcon.net/index.php?id=381> (Accessed February 13, 2009).
15. Workplace Safety & Insurance Board of Ontario. Workwell. Available at: <http://www.wsib.on.ca/wsib/wsibsite.nsf/public/Workwell> (Accessed February 13, 2009).
16. Robson LS, Clarke JA, Cullen K, Bielecky A, Severin C, Bigelow PL, et al. The effectiveness of occupational health and safety management system interventions: A systematic review. *Safety Sci* 2007;45(2):329-53.
17. Bigelow PL, Greenstein SL, Keefe TJ, Gilkey DP. Development of an on-site, behavior-based safety audit for the residential construction industry. *Work* 1998;11:11-20.
18. Laitinen H, Marjamäki M, Paivarinta K. The validity of the TR safety observation method on building construction. *Accid Anal Prev* 1999;31:463-72.
19. Straker L, Burgess-Limerick R, Pollock C, Egeskov R. A randomized and controlled trial of a participative ergonomics intervention to reduce injuries associated with manual tasks: Physical risk and legislative compliance. *Ergonomics* 2004;47(2):166-88.
20. Guldenmund F, Hale A, Goosens L, Betten J, Duijm NJ. The development of an audit technique to assess the quality of safety barrier management. *J Hazard Mater* 2006;130:234-41.
21. Hale AR, Heming BHJ, Smit K, Rodenburg FGTh, Van Leeuwen ND. Evaluating safety in the management of maintenance activities in the chemical process industry. *Safety Sci* 1998;28(1):21-44.
22. Hee DD, Pickrell BD, Bea RG, Roberts KH, Williamson RB. Safety management assessment system (SMAS): A process for identifying and evaluating human and organization factors in marine system operations with field test results. *Reliab Eng Syst Safe* 1999;65:125-40.
23. Hurst NW, Hankin R, Bellamy LJ, Wright MJJ. Auditing – a European perspective. *J Loss Prevent Proc* 1994;7(2):197-200.
24. Papazoglou IA, Bellamy LJ, Hale AR, Aneziris ON, Ale BJM, Post JG, et al. I-risk: Development of an integrated technical and management risk methodology for chemical installations. *J Loss Prevent Proc* 2003;16(6):575-91.
25. Pitblado R, Williams JC, Slater DH. Quantitative assessment of process safety programs. *Plant/Operations Prog* 1990;9(3):169-75.
26. Bigelow PL, Robson LS. Occupational Health and Safety Management Audit Instruments: A Literature Review. Toronto, ON: Institute for Work & Health, 2005.
27. Diekemper RF, Spartz DA. A quantitative and qualitative measurement of industrial safety activities. *ASSE J* 1970;Dec:12-19.
28. Kuusisto A. Safety management systems: Audit tools and reliability of auditing [dissertation]. Tampere, Finland: Tampere University of Technology, 2000.
29. Uusitalo T, Mattila M. Evaluation of industrial safety practices in five industries. In: Mital A (Ed.), *Advances in Industrial Ergonomics and Safety*, vol.1. London: Taylor & Francis, 1989.
30. Eisner HS, Leger JP. The international safety rating system in South African mining. *J Occup Accid* 1988;10:141-60.
31. Eisner HS. Safety rating systems in South African mines. *J Health Safety* 1993;9:25-30.
32. Guastello SJ. Some further evaluations of the International Safety Rating System. *Safety Sci* 1991;14:253-59.
33. Collison JE, Booth RT. An evaluation of two proprietary health and safety auditing systems. *J Health Safety* 1993;9:31-38.
34. Gaunt LD. The Effect of the International Safety Rating System (I.S.R.S.) on Organizational Performance. Georgia: Center for Risk Management and Insurance Research, Georgia State University, 1989.
35. Chaplin R. An evaluation of the use of the International Safety Rating System (ISRS) as intervention to improve the organisation of safety. In: Hale A, Baram, M (Eds.), *Safety Management: The Challenge of Change*. Oxford: Pergamon, 1998.
36. Glendon AI, Boyle AJ, Hewitt DM. Computerized health and safety audit systems. In: Matilla M, Karwowski W (Eds.), *Computer Applications in Ergonomics, Occupational Safety and Health*. Amsterdam: Elsevier, 1992.
37. Dyjack DT. Development and evaluation of an ISO 9000-harmonized occupational health and safety management system [dissertation]. Ann Arbor, MI: University of Michigan, 1996.
38. Dyjack DT, Levine SP, Holtshouser JL, Schork MA. Comparison of AIHA ISO 9001-based occupational health and safety management system guidance document with a manufacturer's occupational health and safety assessment instrument. *Am Ind Hyg Assoc J* 1998;59(6):419-29.
39. Alteren B, Hovden J. The Safety Element Method – a user developed tool for improvement of safety management. *Safety Sci Monitor* 1997;1(3):1-23. Available at: <http://www.monash.edu.au/muar/ippo/issues.html> (Accessed February 24, 2009).
40. Alteren B. Implementation and evaluation of the Safety Element Method at four mining sites. *Safety Sci* 1999;31:231-64.
41. Dyjack DT, Redinger CF, Ridge RS. Health and safety management system audit reliability pilot project. *Am Ind Hyg Assoc J* 2003;64(6):785-91.
42. Redinger CF. Occupational health and safety management system conformity assessment: Development and evaluation of a universal assessment instrument [dissertation]. Ann Arbor, MI: University of Michigan, 1998.
43. Redinger CF, Levine SP. Development and evaluation of the Michigan occupational health and safety management system assessment instrument: A universal OHSMS performance measurement tool. *Am Ind Hyg Assoc J* 1998;59:572-81.
44. Redinger CF, Levine SP, Blotzer MJ, Majewski MP. Evaluation of an occupational health and safety management system performance measurement tool – II: Scoring methods and field study sites. *Am Ind Hyg Assoc J* 2002;63(1):34-40.
45. Redinger CF, Levine SP, Blotzer MJ, Majewski MP. Evaluation of an occupational health and safety management system performance measurement tool – III: Measurement of initiation elements. *Am Ind Hyg Assoc J* 2002;63(1):41-46.
46. Kam YK, Wong YY. Introduction to SMA-CON: Safety audit for construction. In: Zeng Q, Xie X, Wang L, Qian X (Eds.), *Progress in Safety Science & Technology, v 1, Proceedings of the International Symposium on Safety Science and Technology*. Beijing: Science Press, 1998.
47. Schweigert MK, House RA, Holness DL. Occupational health and safety management systems in the Canadian pulp and paper industry: Methods of auditing. *J Occup Environ Med* 1999;41(10):857-62.
48. Fuller C. Benchmarking health and safety performance through company safety competitions. *Benchmarking: an Intl J* 1999;6(4):325-37.
49. Booth RT, Lee TR. The role of human factors and safety culture in safety management. *Proc Instn Mech Engrs* 1995;209(B5):393-400.
50. Pearce W. Club Zero: Implementing OHS management systems in small to medium fabricated metal product companies. In: Pearce W, Gallagher C, Bluff L (Eds.), *Occupational Health & Safety Management Systems. Proceedings of the First National Conference*. Melbourne: Crown Content, 2001.
51. Pearce W. Club Zero: Implementing OHSMS in small to medium fabricated metal product companies. *J Occup Health Safety - Aust NZ* 2002;18(4):347-56.
52. Bunn WB, Pikely DB, Slavin TJ, Parlar S. Health, safety, and productivity in a manufacturing environment. *J Occup Environ Med* 2001;43(1):47-55.
53. LaMontagne AD, Barbeau E, Youngstrom RA, Lewiton M, Stoddard AM, McLellan D, et al. Assessing and intervening on OSH programmes: Effectiveness evaluation of the Wellworks-2 interventions in 15 manufacturing work-sites. *J Occup Environ Med* 2004;61(8):651-60.
54. Teo EAL, Ling FYY. Developing a model to measure the effectiveness of safety management systems of construction sites. *Build Environ* 2006;41:1584-92.
55. Nielsen KJ, Rasmussen K, Glasscock D, Spangenberg S. Changes in safety climate and accidents at two identical manufacturing plants. *Safety Sci* 2008;46(3):440-49.
56. Terwee CB, Bot SD, de Boer MR, Van Der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
57. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments. *Qual Life Res* 2002;11:193-205.
58. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
59. Goutteborge V, Haije Wind H, Kuijer PPFM, Frings-Dresen MHW. Reliability and validity of functional capacity evaluation methods: A systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. *Int Arch Occup Environ Health* 2004;77:527-37.
60. Matheson A, Mughal W, Thomas-Olson L, Spiwak R, Wasdell M. Inter-rater reliability assessment of an OH&S management systems audit tool. Presented at National Occupational Injury Research Symposium, Pittsburgh, PA, October 21-23, 2008.
61. Gillette D, Campbell P, Busby B. The evolution of a radiation safety audit program for a research institution. *Health Phys* 2004;86(2):S80-S84.
62. Birkmire JC, Lay JR, McMahon MC. Keys to effective third-party process safety audits. *J Hazard Mater* 2007;142:574-81.
63. Budworth N, Cox S. Trusting the tools. *Safety Health Pract* 2005;23(7):46-48.
64. Cooper D. Safety management system auditing. In: Cooper D (Ed.), *Improving Safety Culture – A Practical Guide*. Chichester, UK: John Wiley & Sons Ltd., 1998.
65. Henriksson L. Looking for gold. *OHS Canada* 1998;12:48-51.
66. International Organization for Standardization (ISO). ISO 19011: Guidelines for quality and/or environmental management systems auditing. Geneva (Switzerland): ISO, 2002.
67. Karapetrovic S, Willborn W. Quality assurance and effectiveness of audit systems. *Int J Qual Reliability Manage* 2000;17(6):679-703.
68. Grant J, Bricker R, Shiptsova R. Audit quality and professional self-regulation: A social dilemma perspective and laboratory investigation. *Auditing: J Pract Theor* 1996;15(1):142-56.
69. Messier WF, Jr, Kachelmeier SJ, Jensen KL. An experimental assessment of recent professional developments in nonstatistical audit sampling guidance. *Auditing: J Pract Theor* 2001;20(1):81-96.
70. Meyer MJ, Rigsby JT, Boone J. The impact of auditor-client relationships on the reversal of first-time audit qualifications. *Managerial Accounting J* 2007;22(1):53-79.
71. Mohd-Sanusiz P, Mohd-Iskandar T. Audit judgment performance: Assessing the effect of performance incentives, effort and task complexity. *Managing Auditing J* 2007;22(1):34-52.
72. Nelson M, Tan H-T. Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: J Pract Theor* 2005;24(Suppl):41-71.

MEASUREMENT PROPERTIES OF OHS AUDITS

73. O'Leary C. The consistency of individual auditors in performing evaluations. *Managing Auditing J* 2004;19(5):597-605.
74. Pflugrath G, Martinov-Bennie N, Chen L. The impact of codes of ethics and experience on auditor judgments. *Managerial Auditing J* 2007;22(6):566-89.
75. Preuss L. On ethical theory in auditing. *Managerial Auditing J* 1998;13(9):500-8.
76. Richard C. Why an auditor can't be competent and independent: A French case study. *Eur Account Rev* 2006;15(2):153-79.
77. Sweeney B, Pierce B. Management control in audit firms: A qualitative examination. *Account Audit Accountability J* 2004;17(5):779-812.
78. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997;50(1):79-93.
79. Bot SD, Terwee CB, Van Der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: A systematic review of the literature. *Ann Rheum Dis* 2004;4:335-41.
80. Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: A systematic review. *J Occup Rehab* 2007;17(3):504-21.

RÉSUMÉ

Objectif : Les propriétés de mesure des audits de gestion en santé et sécurité du travail (SST) pourraient être importantes pour certaines applications, surtout lorsque les scores d'audit sont considérées comme des mesures de rendement. Nous avons donc voulu répertorier et résumer les résultats de recherche sur les propriétés de mesure (p. ex., fiabilité, validité) des méthodes utilisées pour les audits de gestion en SST.

Méthode : Les bases de données bibliographiques des domaines des affaires, de la médecine et de la SST ont été systématiquement interrogées. Les données probantes de publications pertinentes ont été résumées à l'aide de méthodes classiques de recension narrative des écrits.

Synthèse : Les travaux publiés sur les propriétés de mesure des méthodes utilisées pour les audits de gestion en SST sont rares. Dix-sept méthodes d'audit pertinentes ont été recensées. Nous n'avons pu démontrer la validité de contenu que pour cinq de ces méthodes. La fiabilité inter-évaluateurs n'a été véritablement testée que pour trois méthodes d'audit, et la validité de construit, pour une seule méthode. Il n'y avait aucune étude de fiabilité de test-retest, ni de sensibilité. Les études de fiabilité inter-évaluateurs (cohérence d'un évaluateur à l'autre) ont montré que cette fiabilité est souvent trop faible pour être acceptable.

Conclusion : Il y a des lacunes dans la recherche sur les propriétés de mesure des méthodes de gestion en SST. Les travaux publiés soulèvent des questions quant aux propriétés des méthodes d'audit utilisées actuellement.

Mots clés : prévention des accidents; audit de gestion; santé au travail; reproductibilité des résultats; gestion de la sécurité; études de validation