SAFETY SCIENCE Monitor
ISSUE 1 2004
Article 3
ISSN 1443-8844
VOL 8

WAIT (PART III) – PRELIMINARY VALIDATION STUDIES

CELESTE JACINTO[A, B] AND ELAINE ASPINWALL[A †]

[A] School of Engineering - Mechanical and Manufacturing Engineering, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK [B] Department of Mechanical and Industrial Engineering, FCT/ New University of Lisbon, P-2829-516 Caparica, Portugal

ABSTRACT

This paper is the last of a series of three, all of them associated with WAIT (Work Accidents Investigation Technique), which is a new method for use in the investigation and analysis of occupational accidents. The initial paper (Part I) introduced the WAIT process as well as its theoretical foundations. Part II discussed the results of its application in 56 real accidents. It showed that WAIT fulfilled its intended purpose and provided good coverage in the case of both simple and complex accidents at work, across a variety of industrial sectors.

It is widely accepted that accident data, and the lessons learnt from it, play an important role in the continuous improvement of safety. However, the effectiveness of these improvements relies on the quality of such data, and one must attempt to ascertain not only whether or not a method works (in terms of its applicability and coverage), but also if it works sufficiently well (in terms of providing valid and reliable information). The latter was the objective of the work reported in this paper (Part III). Three studies are discussed, providing a preliminary insight into the technique's validity and reliability. The first is a comparison with another structured technique, the second is an inter-analysts reliability study, and the third, designed for assessing certain other attributes, is based on the subjective opinion of the users of WAIT. The paper concludes with proposals for future work aiming to confirm the (apparently satisfactory) results obtained so far and to study the distribution of causation factors in a larger and statistically representative sample.

1. INTRODUCTION

This paper reports preliminary validation studies on the WAIT technique. Within this research work, *validation* has been defined as the process of examining to what extent the technique is "well-grounded" or "well-founded and fully applicable to the particular matter or circumstances", with the term *valid* being derived from the Latin *valere* = to be strong (Oxford University Press, 2003, *OED-online*). In contrast with the term *testing* (discussed in Part II), validation encompasses the idea of a more formal scrutiny designed to give evidence that something not only works, but also works sufficiently well, i.e., is well founded. The validation of WAIT was carried out through specific studies, designed to measure particular attributes. It provides a more formal approach for accepting or rejecting the hypotheses under consideration, i.e., whether or not particular requirements were satisfactorily complied with, and the degree of compliance.

† Corresponding author: Elaine Aspinwall, Senior Lecturer, School of Engineering, Mechanical and Manufacturing Engineering, The University of Birmingham, Edgbaston, Birmingham B15 – 2TT, UK. Tel: 0044 – 121 – 414 4249 Fax: 0044 – 121 – 414 4152 E-mail: E.Aspinwall@bham.ac.uk

It has been widely recognised that the ability to learn the right lessons from accidents is a key element for the successful improvement of safety. However, the effectiveness of the lessons learnt depends, to a large extent, on the quality of feedback information, and it presupposes that valid and reliable methods are used to obtain accident data. As such, validation studies are fundamental to any technique, and even more important in the case of new ones. This effort may sometimes be considered of little use, since none of the validation approaches is without its limitations. However, the authors believe that an attempt needs to be made to assess, as far as reasonably possible, a method's validity, prior to deciding whether it is worthy of further development.

In practice, the validation of WAIT was carried out by means of two specific studies designed to complement each other in examining the *validity* and the *reliability* of the proposed technique. The first was a comparison with another structured technique and the second an inter-analysts reliability study. A third study is also presented, in which certain other attributes/characteristics of the method were formally assessed through the measurement of its users' subjective opinions.

2. COMPARISON WITH ANOTHER STRUCTURED TECHNIQUE

This study, to assess the *validity* of WAIT, i.e., how "strong" and "well-founded" the method is, was carried out by comparing it with another formal technique. The latter was chosen because it was considered well-established, and was being used in an organisation that has been recognised by the British HSE as an example of best practice (HSE, 1998). The organisation was DuPont Nylon (Gloucester, UK), and the approach they use for *full investigations* of occupational accidents is a fault tree technique, called *Possible Factors Analysis* (PFA). The method is applied for identifying both immediate causes and other causal factors (called *Possible Factors*). The tree technique is performed manually, and all people likely to be involved in an investigation have training on its use. From the findings, the investigation team makes recommendations for each possible factor identified. There is not a checklist for the search: investigators will use their own experience and skills. However, the information contained in the accident reports is, later on, stored in a computerised database, which provides a structured classification scheme for many variables, including *Process Safety Management* (PSM) factors, related to the production process and management of safety.

Procedure used for the comparison

This analysis was of a qualitative, rather than a quantitative, nature as it involved subjective judgement in deciding which factors (or problem-areas) were equivalent, given the fact that the two approaches use different nomenclatures in their classification schemes.

The case used for this comparative study was not an occupational accident, but an accident where only "loss of property" had occurred. It was rather a complex case, with a diversity of failures, thereby allowing a comprehensive study. It had occurred a few days before the study, and both investigations ran almost in parallel, allowing a certain degree of independence at least as far as making recommendations was concerned. A total of 3 sessions (i.e., 3 days, although not consecutive) were necessary to complete this study. Two of DuPont's employees were directly involved in the WAIT analysis: the local Safety Manager and his closest co-worker.

An important consideration is that in this type of procedure (i.e., comparing techniques on the same occurrence) it is virtually impossible to eliminate biases from the analysts, because some of them participate in both processes and they will transport their own ideas from one analysis process to the other. Despite this constraint, comparing a new method with others is common in validation studies, but care must be taken when drawing conclusions. As a safety researcher claimed (Dr. Inge Svedung, *personal communication, London, October 2002*), even if all goes well, one can only conclude that the new method probably is *at least as good as* the other, but not necessarily better. This is why a study of this nature needs to be complemented with others, such as, for instance, an inter-analysts reliability study – which is also presented in this paper.

3. RESULTS AND DISCUSSION

The initial part of the WAIT process is likely to promote good adherence to "reality" because it focuses on "observable" elements. In terms of assessing validity, therefore, the major concern was at the level of the inferred findings. Only organisational and management weaknesses will be compared here, since the PFA technique does not include cognitive aspects in its analysis. The comparison was performed by looking at similar "problem-areas" identified by each approach. Table 1 summarises the results.

Table 1- Comparison between WAIT and PFA techniques

(organisational and management factors)

	problem-areas	WAIT	PFA
1	Contractors: safety & performance; management of,	?	? *
2	Training problems - people and/or methods	?	? *
3	Technical controls and barriers	?	?
4	Design of plant and/or equipment	?	?
5	Work procedures and methods (includes test methods)	?	?
6	Permits-to-work and supervision	?	?
7	Hiring & placing: promotions and staffing competence	?	O
8	Communication practices and methods	?	O
9	Risk assessment	?	?

* Contractors and Training were the only two registered in the database of PSMs, but from the investigation report, one could see that other areas were identified and reported - therefore they should be part of this comparison.

From the table, it becomes apparent that there were two particular areas addressed only by WAIT: "hiring & placing" and "communications". The first was associated with a person who had recently been promoted but was not fully aware of his new duties and roles. The latter was identified on two particular occasions, but the most relevant was associated with the use of different measurement units on verbal communications between different individuals (bar *versus* psi, in this case).

Since "recommendations" are an important output of any method, these were also compared. Coincidently, each method produced 9 recommendations and they were, in the main, quite similar, addressing the problems identified. There were two slight differences, which "mirror" the differences identified in Table 1. Recommendations produced by WAIT highlighted the need for using uniform measurement units for reducing errors. They also addressed the issues of supervision and promotions. On the other hand, sharing the learning, not only at site level but also with other sites, was highlighted in DuPont's recommendations. In WAIT, however, this is an "automatic" and intrinsic part of the process, as is explained in its user's manual.

In this one validation study, therefore, it appears that WAIT promotes a more comprehensive/detailed scrutiny. This apparent advantage of WAIT, that needs confirmation, is associated with its classification schemes, which can be used as "check lists". They do not exist in the PFA technique where investigators rely mainly on their own skills and experience.

For discussing validity, other aspects might be taken into account. In contrast with predictive methods, WAIT is simply a *retrospective analysis* technique. There is no need to make predictions: the occurrence under analysis has already happened and all elements are there. What is of great importance here, is how comprehensive is the search, and whether or not all relevant elements will actually be found. From the overall results so far, including those of the testing stage, it seems that one of WAIT's strengths is its classification schemes; not only are they comprehensive but, as has already been mentioned, they can also be used as "check lists" during the search, and prompt analysts to scrutinise a variety of possible weaknesses. Naturally, this is not a guarantee that all relevant problems will be identified, but it can be argued that those which are, are likely to be valid. However, this study was based on a single case - although complex - and it is still too early to draw final conclusions. It can only be said that WAIT seems to be well-grounded and relatively strong in providing a means for a thorough investigation.

In pragmatic terms, perhaps the most relevant outcome of this experiment was the fact that the Safety Manager of the site concerned found it worthwhile to make a presentation of WAIT to the DuPont European General Safety and Contractor Safety Network (in Bad Homburg, September 2002). The presentation was well received and the delegates decided that a small team made up from people representing sites in the UK, Holland, Spain and Germany would trial the use of WAIT alongside their normal processes and feedback to the network (annual meetings). Apparently, the most appreciated characteristics of WAIT were the standard questionnaire provided and the classification schemes (Kevin Edwards, *e-mail, 26th September 2002*).

4. INTER-ANALYSTS RELIABILITY

The aim of this second study was to examine the degree of agreement between results produced by different users, i.e., how consistently different people use the method.

The first stage of the WAIT process (basic investigation) focuses mainly on factual information and observable elements; as such, it is less susceptible to bias from the analysts. In contrast, the second stage (in-depth analysis) requires a certain amount of inference, and the user's manual strongly recommends the use of a *team* for performing the analysis: this is where differences in judgement are more likely to affect the final results – and naturally, the applicable recommendations. Bearing this in mind, the study was carried out by comparing the results produced by two *teams*. In practical terms the inter-analysts *reliability* was measured, at different stages, in the following way:

Inter-analysts reliability: "the number of common factors found divided by the total number of factors (within the same category) identified by all teams, as a %".

Procedure of the reliability study

The work was divided into two different sessions: one to give basic training to all participants, and a second for analysing and comparing the results of two cases under consideration. There were a total of 6 subjects, all from Land Rover (Solihull, UK). Each team was made up of 3 members with the following identical composition:

One team leader (a senior safety person, with a NEBOSH qualification and several years of experience; these were the local site Safety Manager and the Group's Safety Manager)

Two Safety Representatives (with formal training and some experience in accident investigation)

Training session - on the first day all participants received basic training in the use of WAIT, lasting approximately 3 hours. Since people seem to retain more of what they do, rather than of what they hear or read, half of this session focused on the practical application of the technique, in which three accidents were analysed in some detail.

Reliability study - this took place on the second day, at the University of Birmingham. On this occasion, the authors acted mainly as facilitators and timed the activities. The session lasted all day and two different accidents were analysed: one in the morning and another in the afternoon. Two main criteria were used in the selection of the cases:

1- they should be neither over-simple, nor too complex, and

2- to avoid pre-judgement biases, they should be *unknown* to all subjects (i.e., no Land Rover cases). In addition, the relevant information that needed to be given should be objective, i.e., as factual as possible, and allow sufficient understanding of the context without visiting the concerned site.

A few cases were selected, but time restrictions only allowed the analysis of two. As will be shown, however, the selection was not completely successful and the second case (case B) proved to be a "poor case for study" due to lack of relevant information for the participants. Only too late, it was realised that photographs of the accident scenario would have been extremely useful in this case, but they were not available at the time.

Basic procedure – the working procedure was similar in both cases analysed. At the outset of each one of them, the participants received a brief description of the accident under consideration and the answers of the people involved to the WAIT standard questionnaire. To enable the search of other (less obvious) problems, they were also given, in writing, some additional information. This was factual information and, as far as possible, free from pre-judgements (e.g.: if there was a previous risk assessment and the information it contained, or whether there was a written procedure for a certain task, or if there was an established routine for something - even if it was verbal, etc.). The first case studied (A) involved a worker whose arm was crushed by a moving part of a handling truck (a photograph of the truck was available); the second case (B) involved a slip/fall of a worker who was transporting an object by hand.

Once the analyses were completed, the results obtained by both teams were compared – this was done not only between the two teams, but also with the original tables produced on site – and discussed.

Results and discussion of reliability

The final results of this reliability study are shown in Table 2. The first case analysed (A) did not pose major difficulties and the participants were able to carry out their work without being provided with more information. The inter-analysts reliabilities obtained in this case ranged from a 60% to a 100% match (~ 77% average) suggesting that the overall agreement was relatively high.

Table 2- Summary results of the inter-analysts reliability study

Factors analysed	Team 1	Team 2	On site	Total *	Agreement	Reliability
	Number of factors identified				Common **	% of agreement
Case A						
WPC (workplace influencing factors)	4	3	3	4	3	75%
IND (individual factors)	1	1	1	1	1	100%
JOB (job/task & technology)	7	7	5	7	5	71%
O&M (organisational & management)	3	3	5	5	3	60%
Case B						
WPC (workplace influencing factors)	4	4	3	5	1	20%
IND (individual factors)	1	1	1	1	1	100%
JOB (job/task & technology)	3	4	2	6	1	17%
O&M (organisational & management)	1	3	2	5	0	0%

**** total identified by all teams (some are common) /// ** only those that were common to all teams***

It is difficult to establish what percentage should be considered a "good result", for there are no reference values of other identical studies against which to compare. Nothing else but common sense suggests that less than 50% would probably be "poor" and a very high percentage would be "suspicious" (i.e., probably biased) considering that findings depend, to a large extent, on individual judgements. A validation study described by Hollnagel (2000), using two independent analysts, reported a "match percentage" ranging from 42% to 100%, with an average of 67.8% - and this percentage of agreement had been considered reasonably good. The method being assessed in the referred study, however, was used in a *predictive analysis* and it addressed human performance. There are no grounds for stating that a value of the same magnitude could also be considered good for a *retrospective method*, as is the case with WAIT. Quite recently, another study (Wright and Van der Schaaf, 2003) using two separate analysts, reports around 78%-80% for the inter-assessors reliability of the method CIRAS (Confidential Incident Reporting and Analysis System). This is the UK railway national system for the reporting and analysis of railway near-misses. The reporting is voluntarily and confidential, and the investigation is mainly based on *an interview* with the concerned driver. In contrast with more structured approaches, this is a relatively simple investigation process, which may facilitate a good level of agreement between different assessors. Finally, both studies found in the literature used two analysts, whereas in this research work the comparison was made between three groups, which increases the degree of difficulty, i.e., the "matching" or overlap of results is more likely to be reduced.

The second case that was analysed was disappointing and Table 2 shows the very low agreement obtained (~34% average match). This was a badly selected case-study because the additional information provided to the participants was not sufficient – a fact that had not been perceived *a priori*. On this occasion, both teams stated that they needed to know many more details, but the authors felt that, giving them more would influence the analysis and the results would be biased. Therefore the work continued in an attempt to see what would come out just using the information provided. During the final discussion, it was proposed that this case be discarded as "not good for study", but the deliberations proved to be fruitful and it was felt that ignoring this analysis would be a missed opportunity.

The main lesson learnt from this second case-study was that, in similar situations (i.e., performing the analysis away from the accident scene/place), it is highly advisable to use photographs in addition to written information. They facilitate a much better understanding of the situation, and offer little chance for bias. It was found, however, that there was a relatively good match between the findings of Team 1 and the original investigation on site (more than 50% agreement in all categories), but when the findings of Team 2 were considered, the number of matches was drastically reduced, i.e., there were less "common factors" to all three teams. Team 2 followed a different direction from the outset as they focused their attention on the object being carried (by hand) by the injured person. A simple photograph of the object being transported (and the bench where the work was being done) would, probably, have been enough to understand that this was a false “problem”. In the first case (A) there was a photograph of the handling truck involved in the accident and the participants considered this very useful. Other ways of providing unbiased information is to have copies of relevant original documents, but this can prove to be very difficult to achieve, because many firms want to remain anonymous and will not allow photocopies of their documents to be used outside the firm. Naturally, the “ideal” approach would be having independent analysts (or teams) carrying out the investigation, without restrictions, on site and in parallel.

The discussion of case B also revealed that, although there was poor agreement *between* teams, the factors registered were consistent and logical *within* the same team. This may indicate that the classification schemes for all categories are comprehensive enough and allow the establishment of logical and coherent links throughout the whole process.

In spite of the problems reported, both cases under scrutiny obtained a good level of agreement (almost 100% match) in what had been registered as “active failures” and in the coding of applicable variables. For the reasons discussed above, case B is not good enough to provide values (or to enable conclusions to be drawn) on the percentage of agreement. However, it was decided to maintain it as part of this study, to help prevent others from repeating the same “wrongs” in similar situations.

This preliminary evaluation of the inter-analysts *reliability* of WAIT is, therefore, mainly based on the results of a single case-study (case A). The average agreement value obtained (~77%) seems to be reasonably good, and should be confirmed with more studies, preferably carried out *in situ* by independent researchers or analysts.

5. USERS’ JUDGEMENT OF WAIT ATTRIBUTES

This final Section examines how 5 particular attributes/characteristics of WAIT were perceived by the targeted users and to what extent they satisfied the initial pre-requisites. The 5 attributes under scrutiny were:

? *Usefulness* in improving safety and its management (through appropriate recommendations and review of risk assessment)

? *Simplicity* of use (by common safety professionals in their organisations)

? *Auditability* (in terms of providing auditable analyses and retrieval of information)

? *Sensitivity to the context* (for both workplace influencing factors and organisational/management conditions)

? *Storage of data* (in terms of usefulness of the coding structure and format of the data)

Procedure used for measuring users’ opinion

During the pilot run (see Part II), the authors gained an insight into the users’ opinions on many of the aspects discussed here. However, this was not formally measured at the time, since the main concern was to examine applicability and coverage.

For the second run, therefore, a procedure was designed to assess these attributes in a systematic way and as objectively as possible, although based on the users’ subjective judgement. A questionnaire was produced for this purpose, the questions being associated with a rating scale of 1–7.

The users in the companies involved in this study (13 users in all) were asked to give their opinion by scoring each attribute. Many also gave suggestions for improving the technique.

Results and discussion of users’ judgement

Table 3 summarises the users’ scores. The last three columns of the table show statistical results assuming a *metric interval scale*. In the last column, the average (mean score) is converted to a percentage scale (0-100%), with which most people are more familiar.

Table 3 - Users' judgement (scores given to particular attributes)

scores →	1	2	3	4	5	6	7	Average	Std.	Average
Attributes	**? of users giving the scores**							**[1-7]**	**Deviation**	**[% scale]**
1. Usefulness in improving safety										
1.A. through recommendations			1	1	4	7		*5,31*	*± 0,91*	*72%*
1.B. through risk assessment					5	3	5	*6,00*	*± 0,88*	*83%*
2. Practicability / simplicity of use			2	2	3	4	2	*5,15*	*± 1,29*	*69%*
3. Auditability	1			1	4	5	2	*5,31*	*± 1,49*	*72%*
Auditability (after deleting the outlier)								*5,67*	*± 0,85*	*78%*
4. Sensitivity to context										
4.A. Influencing factors				2	2	6	3	*5,77*	*± 0,97*	*80%*
4.B. Organisational conditions				1	3	6	3	*5,85*	*± 0,86*	*81%*
5. Storage of data				2	1	4	2	*5,67*	*± 1,05*	*78%*

scores of 13 users (only 9 people scored attribute 5)

Statistical results assuming a ***metric scale (equal intervals)***

Although applying metric statistics is common practice in research work using rated questionnaires, this is not always the most appropriate way of analysing this type of data. Rated scales are, typically, *non-metric ordinal* scales, and the correct statistical indicators are the median, frequency and quartiles, rather than the average and standard deviation (Kinnear and Taylor, 1983; Hair *et al*, 1998). However, if one assumes that the differences between any two consecutive points of the scale are equal, then the results can be treated as metric data from an interval scale (Tabachnick and Fidell, 1989). To facilitate the analysis, this kind of assumption was made in this study, and the metric statistics in Table 3 will be used for the discussion.

In general, all characteristics were scored highly with an average of 69% – 83%. The best attributes were perceived to be *usefulness for risk assessment* (1.B = 83%) and *sensitivity to the context*; the latter including both *organisational conditions* (4.B = 81%) and workplace *influencing factors* (4.A = 80%). Not only did these gain the highest average score, but they also had a low standard deviation, showing a high level of agreement between the users.

The high agreement is particularly encouraging if one considers that the users involved ranged from highly qualified safety personnel to those with little training in safety matters, who carry out this function on a part-time basis, together with their main job (e.g.: Production Manager, Sales Manager, or General Manager).

It appears then, that usefulness in addressing risk assessment and sensitivity to capture the context (and identify its weaknesses), are seen, by a variety of users, as very strong characteristics of the method.

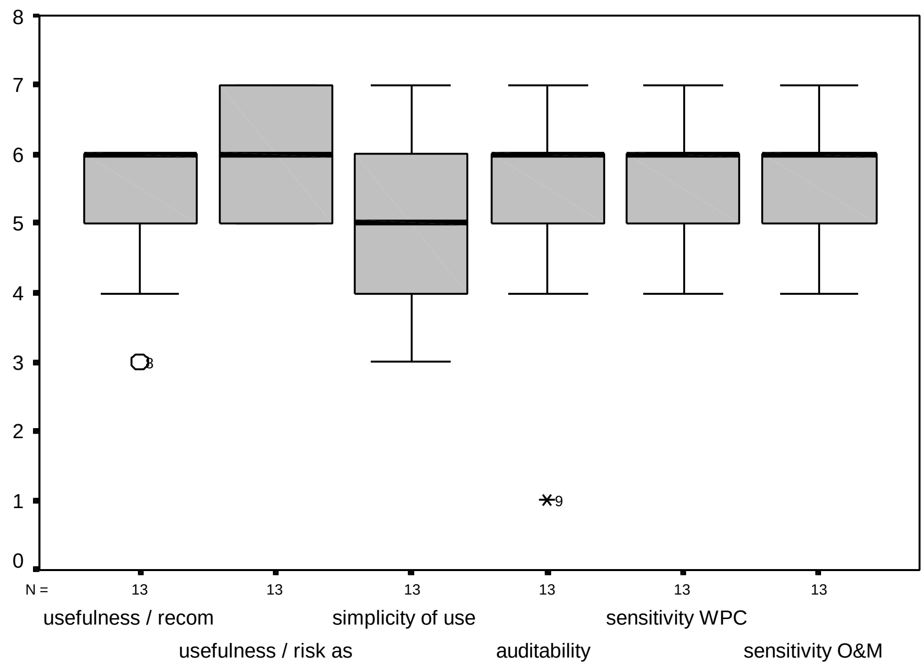
In contrast, lower scores were given to *simplicity* (attribute 2 = 69%), immediately followed by *usefulness in suggesting and ranking recommendations* (1.A = 72%) and *auditability* (attribute 3 = 72%). Additionally, the standard deviations show that, in these cases, opinions varied considerably. These three attributes were examined in more detail as possible candidates for improving WAIT.

When individual data was analysed, it was found that people from the "safety-minded firms" were usually more critical; these were commonly, but not necessarily, the large organisations. The exception was *simplicity*, where the opposite trend was observed. Not surprisingly, users from large firms found the approach practical and quite simple to use, whereas those of small firms found it less simple to apply. In spite of this, they also stated that they felt that WAIT would be easier to use after gaining more practical experience, and if they had more time available for safety-related activities.

With regard to the attribute *usefulness/recommendations* (1.A), opinions varied from 3 to 6 points, although 7 users rated this with 6 points; thus 6 is the statistical mode and represents the majority. Of the total (13), only 2 users gave a low scoring, which impacted the average. Both gave the same reason: in their opinion, WAIT was very useful in suggesting appropriate recommendations, but poor in terms of ranking them. Land Rover, in particular, considers that the method's manual did not provide enough guidance on how to present and prioritise recommendations. Their suggestion was to create a specific form for registering recommendations in two categories: immediate and long-term. The Safety Manager of the Engineering Directorate of QinetiQ also gave a similar suggestion. A form of this type is now included in the final version of the user's manual.

Auditability was a very particular case: Table 3 shows that one user rated this as the lowest possible (1 point). A graph plot of the results, in Fig.1, shows that this particular case was, from a statistical point of view, an *outlier*[1] and therefore, should be excluded. In fact, if the outlier is deleted, the resultant average is 5.67 ± 0.85 (~78%), which makes this attribute as good as the best scored. The viewpoint of the person concerned was that each single document, including the WAIT manual and its analysis tables, should state a date/period for maintaining the records (~40 years would be his recommendation).

Figure 1 – Box Plot of the User's Scores



(Note:the attribute "storage of data" is not plotted in the chart, because the number of users was different: four people did not feel confident enough to judge this characteristic)

The authors agree that this is an important requirement, but is independent of the method or technique used. Such a requirement needs to be explicitly stated in a safety management procedure; it concerns not only accidents, but also all other types of records (such as health surveillance, for instance). In addition, this issue of record keeping is becoming a legal concern at international level – for all organisations, regardless of their size, activity and methods used. The ILO (2002) recently produced a provisional *protocol* addressing the *recording and notification of occupational accidents*, in which [Art. 3,c] the official procedures for recording shall determine "…*the duration for maintaining these records*". If nothing else, in the light of these new developments, the issue of record keeping is clearly separated from the methods used.

Finally, *storage of data* (attribute 5 = 78%) needs an explanation. From this result, it appears, at first sight, that this is a good attribute of WAIT – but the case is not quite so for the majority of users. When answering this question the users actually seemed to believe that the proposed structure for *coding data* would be very useful in many instances, namely to allow statistical studies on accident causation and the development of better prevention measures. But… and it was a big BUT: this was assuming that the coding is done by authority officials (i.e., by civil servants using text information from the official notification forms).

Most considered the coding of information as time-consuming and unnecessary for themselves, and therefore not worthwhile. In their opinion, constructing the analysis table using their own words would be more than sufficient and adequate. Only a couple of very large organisations appreciated the coding structure as a virtue of WAIT. Not surprisingly, these organisations had full-time personnel allocated to safety, and frequently had

---

[1] Outliers are cases with such extreme values (on one variable) that they unduly influence statistics; these cases seem to be unattached to the rest of the distribution (Tabachnick and Fidell, 1989; p.66)

some difficulties in comparing statistical results between their many different sites. For these, using a uniform coding system, and incorporating the "official European variables", was an attractive idea.

Based on the above (quite understandable) criticism, the user's manual was amended. The coding of information has become "optional" instead of being an intrinsic part of the process, which seems a more pragmatic approach for making WAIT more attractive to all kinds of enterprises.

Finally, Table 4 summarises the main advantages and limitations of WAIT, giving a general overview of the most relevant findings of the experimental work reported in both papers: Part II and Part III.

Table 4- Main advantages and limitations of WAIT

Advantages	The investigation process is systematic and relatively simple to perform. A User's Manual is provided.
	Specifically developed for use in occupational accidents.
	Provides a two-stage investigation, making good use of available resources.
	Takes into account all kinds of causation factors, both immediate and remote.
	Provides good coverage in many work activities, and particularly in industrial environments.
	The standard questionnaire provided for gathering information is considered very useful by users.
	Relates, in a systematic manner, accident investigation (reactive monitoring) with risk assessment (proactive monitoring).
	Has the ability of identifying management and organisational weaknesses even in companies which have no formal H&S management system.
	Directs investigators to make recommendations and to review the management of safety.
	Introduces the concept of *positive influencing factors* – or success criteria – for searching for alternative ways of learning lessons and devising preventive measures.
	Makes a clear distinction between observed facts and inferred findings.
	Validity and reliability are likely to be strong, especially in the first stage (basic investigation process).
	Has the potential to facilitate the harmonisation process in the EU.
Limitations	The classification schemes limit the use of the technique to industrial settings.
	There is no objective criterion for deciding which active failures to include in the analysis.
	Not well suited for activities were people work alone (e.g.: self-employed, family workers, farmers).
	Coding of information is time-consuming, particularly for the active failures.
	Lack of maturity.
	Inter-assessors reliability needs confirmation from independent studies.
	Its use requires some training: at least 3-4 hours are recommended.

6. CONCLUSIONS

This final paper, in the series of three, has discussed the results of preliminary validation studies. The first was a comparison between the results of WAIT and another structured method (a tree technique) being successfully used by DuPont (Gloucester) in the investigation of occupational accidents. The second was an inter-analysts reliability study, in which the results obtained by two different teams, both from Land Rover (Solihull) were compared. Both studies were designed to complement each other in examining the validity and reliability of WAIT. As was discussed, these kinds of validation approaches, although quite common in research work, have limitations and care should be taken when drawing conclusions. Despite this, when one takes into account the results of both studies, it seems that WAIT is relatively "strong" - or well founded - and the results suggest that this new technique is worthy of further improvement and validation.

Finally, certain attributes of the proposed technique were assessed through a quantitative measurement based on its users' judgement. In total, five characteristics were examined using a set of questions associated with a rated scale. In these users' opinion, the best attributes of WAIT were its ability to address risk assessment and its sensitivity to capture the context (i.e., usefulness in identifying weaknesses at all levels). Auditability was also seen as a very strong characteristic of the method. In contrast, two attributes were not so highly scored: simplicity of use and storage of data. However, the opinions varied considerably, and in both cases, this was associated with the comprehensiveness of the coding structure. For some (usually the large organisations) the coding of information was seen as a strong advantage, whereas for others it was deemed unnecessary and time-consuming.

This experimental work and the many people involved in it, provided ideas and useful suggestions for refining the method further. As a result, certain alterations have already been incorporated into the method's User's Manual for future empirical work.

As with any method, WAIT has its limitations; these were summarised in Table 4. Its classification schemes restrict the use of the technique to industrial environments and are not well suited for activities where people work alone, or in rather small groups, such as the self-employed, family workers or farmers. Overall, however, it seems that WAIT is a promising tool although this is still early days and more needs to be done to confirm the findings to date.

Proposals for future work

The experience gained during the experimental work allowed the outline of a future research agenda addressing three main questions:

1. To what extent will the harmonised variable "deviation" actually affect the reliability of official recording systems?

2. How can the validity and reliability of WAIT be confirmed in a more independent manner? And to what extent will the results obtained improve with experience and more training?

3. What will the distributions of all causal factors look like if a statistically representative sample is used?

For answering the first question, two pre-conditions need to be satisfied: (1) the new harmonised system needs to be in place, otherwise performing a study is not worthwhile, and (2) the authorities (of at least one country) need to be involved, i.e., this needs to be a joint-project. Answering the other two questions seems viable in a relatively shorter timeframe. However, a delicate balance needs to be struck: some backup from the enforcement authorities is essential to facilitate timely access to organisations, but researchers should be allowed to keep their findings anonymous for ensuring independence and the full co-operation of all involved. On the other hand, given the huge amount of reported accidents per year, a pragmatic approach would need to restrict the study to a particular sector of activity (or a particular region) to be able to handle a statistically representative sample.

ACKNOWLEDGEMENTS

This paper is part of a research programme leading to a PhD thesis, which is being co-financed by "Universidade Nova de Lisboa" and "Fundação para a Ciência e a Tecnologia" (PRAXIS XXI – SFRH/BD/817/2000), Portugal. The research work is being carried out at the School of Engineering - Mechanical and Manufacturing Engineering, University of Birmingham, UK.

The authors are immensely grateful to DuPont Nylon (Gloucester) and Land Rover (Solihull) for their kind co-operation in the validation of WAIT. This gratitude is extended to the 13 individual users from many other firms who gave their opinion and provided useful suggestions for further improvement.

REFERENCES

Hair, J.; Anderson, R.; Tatham, R. and Black, W. (1998). Multivariate Data Analysis. 5th Edition. Prentice Hall, New Jersey.

Hollnagel, Erik (2000). Looking for errors of omission and commission or The Hunting of the Snark revisited. *Reliability Engineering & System Safety*, 68, 2000, pp. 135-145.

HSE (1998). Managing Risk – Adding Value. How big firms manage contractual relations to reduce risk. The British Health and Safety Executive, HSE Books.

International Labour Organisation - ILO (2002). Protocol on recording and notification of occupational accidents and diseases. ILO, June 2002, Geneva, International Labour Office

(full document available at: http://www.ilo.org/public/english/standards/relm/ilc/ilc90/comreps.htm - Recording and notification of occupational accidents and diseases)

Kinnear, T.C. and Taylor, J.R. (1983). Marketing research: an applied approach. 2nd Edition. McGraw-Hill, New York, London.

Oxford University Press (2003). Oxford English Dictionary – OED online. O*nline version* with updates of the 2nd Edition 1989, 20vols. Oxford University Press. Available on the Internet at: http://www.oed.com (for subscribers only)

Tabachnick, B.G. and Fidell, L.S. (1989). Using Multivariate Statistics. 2nd Edition, Harper Collins, New York, London.

Wright, L. and Van der Schaaf, T. (2003). Accident versus Near-Miss Causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. *Proceedings* of the JRC/ESReDA Seminar on Safety Investigation of Accidents. Petten, The Netherlands, 12-13 May 2003, pp.234-240