

Four studies, two methods, one accident – An examination of the reliability and validity of Accimap and STAMP for accident analysis



Anastacio Pinto Goncalves Filho^a, Gyuchan Thomas Jun^{b,*}, Patrick Waterson^b

^a Ministry of Labour, Federal University of Bahia, Brazil

^b Human Factors and Complex Systems Research Group, Loughborough Design School, Loughborough University, Loughborough LE11 3TU, United Kingdom

ARTICLE INFO

Keywords:

Accident analysis
Accimap
STAMP
Reliability
Validity

ABSTRACT

The validity and reliability of human factors and safety science methods are some of the important criteria for judging their appropriateness and utility for accident analysis, however these are rarely assessed. The aim of this study is to take a closer look at the validity and reliability of two systemic accident analysis methods (Accimap and STAMP) by comparing the results of four studies which analysed the same accident (the South Korea Sewol Ferry accident) using two methods. Studies 1 and 2 used Accimap whilst Studies 3 and 4 applied STAMP. The four studies were compared in terms of analysis procedure taken, level of detail, causal factors identified, and the recommendations for improvements suggested by the methods. The results of the causal factor comparison indicate that the reliability (degree of overlap of causal factors identified from the same method, i.e. inter-analyst overlap) of STAMP (65%) is higher than Accimap (38%). The validity (degree of overlap of causal factors identified from two different methods) is as low as 8%. The comparison of recommendations indicates that STAMP-based analyses produce a wider range of recommendations across multiple system levels while Accimap-based analyses tend to focus on whole system-related recommendations. These findings suggest that the use of a more structured method like STAMP can help produce a more reliable accident analysis results.

1. Introduction

A variety of systemic accident analysis methods have been developed since the 1990s, partly as a way of responding to the increasing complexity of socio-technical systems across a range of domains including healthcare, nuclear power, rail and marine transportation (Waterson et al., 2015, 2017). These systemic accident analysis (SAA) methods draw on sociotechnical systems and control theory (e.g., Accimaps - Rasmussen (1997), STAMP - Leveson (2004)) or resilience engineering (e.g., FRAM - Hollnagel (2004)). They commonly illustrate the diversity of causal factors across different levels of the systems, their interactions and the role played by external influences such as political, cultural, financial, and technological circumstances (Branford, 2011). The extent to which methods for systemic accident analysis produce outcomes which are valid (e.g., the degree to which the accident analysis method successfully identifies the causes of an accident) and reliable (e.g., the degree to which accident analysts using the same accident analysis method produce similar causal representations) are often viewed as an important criterion for judging their appropriateness for accident analysis (Waterson et al., 2017). According to Underwood and Waterson (2013), a lack of validation is the key issue which may

influence the use of the system approach by practitioners. Likewise, Jacinto and Aspinwal (2004) and Stanton and Young (1999, 2003) argue that the ability to learn the right lessons from accidents and to use them with confidence is dependent upon valid and reliable methods. As such, reliability and validation studies of accident analysis methods are widely viewed as desirable and a prerequisite for their use (Ryan, 2015).

There have been a few attempts to conduct various forms of comparative studies of Accimap and STAMP (Johnson and Almeida, 2008; Katsakiori et al., 2009). Salmon et al. (2012) conducted a self-reflective comparative study of Accimap, HFACS and STAMP based analysis of the Mangatepopo gorge incident carried out by different experts. Underwood and Waterson (2014) evaluated the ATSB, Accimap and STAMP methods using two criteria: (1) coverage of systems theory concepts and (2) usage characteristics (e.g., reliability and validity). Their findings, in combination with others (e.g., Waterson et al., 2017) point to a need for a more systematic assessment of the validity and reliability of accident analysis methods. The aim of the current study is, therefore, to systematically assess the validity and reliability of two systemic accident methods: Accimap and STAMP. We identified four studies which analysed the same accident using Accimap and STAMP

* Corresponding author.

E-mail address: G.Jun@lboro.ac.uk (G.T. Jun).

methods. Two studies applied Accimap (Kee et al., 2016; Lee et al., 2017), while of the other two applied STAMP (T. Kim et al., 2016; Kwon, 2016). Three studies were published in peer-reviewed journals (Kee et al., 2016; T. Kim et al., 2016; Lee et al., 2017) and the final one was an MSc dissertation (Kwon, 2016). In what follows, we first describe how the validity and reliability of accident analysis methods have been defined and evaluated in the literature (Section 2). Section 3 describes the criteria and procedure we used in order to assess reliability and validity across the four studies. The results, both qualitative and quantitative, are presented in Section 4. A final section (Section 5) discusses our findings and considers their implications for systemic accident analysis.

2. Validity and reliability of methods for systemic accident analysis (SAA)

2.1. Validity

According to Branford (2007) the validity of accident analysis methods can be considered from two perspectives. One relates to the validity of the method itself that is, the extent to which the method is designed in a way that does what it is intended to do. The focus, in this case, is whether the process underlying the method is appropriate for its intended purposes. Both Accimap and STAMP methods, in theory, are valid from the first perspective, because they are developed to provide insights into how and why an organisational accident occurred throughout the sociotechnical system. The second way of considering the validity is to focus not on the validity of the method itself, but on the validity of the results obtained when the method is used. In this case, the question of whether or not a method does what it is intended to do focuses, not on whether the method and process itself is appropriate for producing the required outcomes, but on whether or not the results obtained through the use of the method are, in practice, those that it is intended to produce. This is referred to as empirical validity - “the degree to which the method works with real cases in a real sample” (Branford, 2007). The two perspectives on validity are clearly related since, if the outcomes of a method are valid, there is an implication that the method used to obtain them is also valid. In fact, some analysts do not distinguish between these perspectives at all, arguing that, “To speak of the validity of a method is simply a shorthand way of referring to the validity of the data or accounts derived from that method” (Maxwell, 1992). However, unlike the validity of a method, the validity of its results can be examined empirically.

Branford (2007) proposed the following four approaches to evaluating the validity of accident analysis methods:

- Evaluations of results against objective external criteria, for example, when a 'gold standard' (the results of a previously validated method) is available and the validity of the results is clearly evident by their agreement or disagreement with this standard. This approach, however, is difficult to use because a 'gold standard' is rarely available;
- Evaluations of results in terms of their internal logic. This approach has limitations that results in terms of their internal logic do not necessarily ensure that the correct answers are reached;
- Evaluation of the results against those of experts. This approach would be applicable only when there is an accident analysis carried out by 'experts'; and,
- Evaluations of the degree of similarity between results obtained from different methods. In this study, the last approach, although there is a potential challenge of comparing results in different format and wording, is used to evaluate the degree of similarity between results obtained from Accimap and STAMP.

2.2. Reliability

Reliability concerns the consistency or repeatability of a method's results (Kirwan, 1992). In order to be reliable, methods must produce outcomes that are “independent of the measuring event, instrument, or person” (Kassarjian, 1977). If the method's outcomes do not vary as a result of these factors, confidence can be gained that the method provides “a source of consistent information” (Militello and Hutton, 1998). According to Goode et al. (2017), the type of accident analyses methods may impact upon reliability. For example, methods which provide a classification scheme of contributing factors to guide the analysis (e.g., Human Factors Analysis Classification System - HFACS), are likely to be more reliable relative to those that do not provide a taxonomy or set of categories for contributory factors (e.g., Accimap and STAMP - Salmon et al., 2012). Hale et al. (1998) argued that the reliability of results can be affected by not only the type of method, but also the background of the analyst such as their education, training, knowledge and experience.

The reliability of accident analysis methods can be evaluated by two approaches. First, the focus of the evaluation can be on the intra-analyst agreement, the extent to which a single analyst using the method produces similar results at different times or when analysing different cases which have similar characteristics. Alternatively, the focus can be on the inter-analyst agreement, the extent to which the outcomes are consistent regardless of the analyst (Pounds and Isaac, 2003). Studies into the reliability of safety-related analysis techniques, such as those designed to analyse and classify human error and human factors issues, tend to take the latter approach, focusing on the level of agreement obtained by independent analysts studying the same data (e.g., Gordon, Flin and Mearns, 2005; Isaac et al., 2003; Johnson and Holloway, 2003). In this study, reliability based on the inter-analyst agreement is assessed specifically on the extent to which different analysts, who access to the similar source data, identify the same causes and produce the same recommendations in their accident analyses.

3. Procedure

3.1. A Brief description of the Sewol Ferry accident

On April 16, 2014, Sewol Ferry, the South Korean ship carrying 476 passengers from Incheon to Jeju Island, sank disastrously. The 18-year-old Japanese-built ship was purchased by a company named Chonghaejin, which added two more floors to the ship to hold more passengers, making the ship extremely unstable. During the voyage, when the ship made a sharp turn, lost its balance and started to list. When Captain Jun Seok Lee communicated with the Vessel Traffic Service (VTS) for help, the Captain made questionable decisions such as telling the VTS that the passengers could not evacuate and instructing the passengers to stay on-board. By the time the captain finally told everyone to evacuate, it was too late. As the result of this accident, 304 people, mostly high school students, lost their lives in what is considered to be one of the most tragic maritime accidents in the history of South Korea (Jun, 2015; Lee et al., 2017).

3.2. Systemic accident analysis methods: Accimap and STAMP

Two systemic accident analysis methods (Accimap and STAMP) were selected in this study since they are well established and currently dominant in human factors and system safety literature (Zhou et al., 2018; Nayak and Waterson, 2016; Stevens and Salmon, 2016; Schmid et al., 2018; Düzgün and Leveson, 2018; Valdez Banda and Goerlandt, 2018; Canham et al., 2018).

Accimaps were developed by Rasmussen (1997) and has subsequently been used to analyse accidents in various domains, e.g., aerospace (Johnson and de Almeida, 2008), patient safety (Waterson, 2009), outdoor activities (Salmon et al., 2012) and marine

Table 1
Overall comparison of the four studies.

Category	Accimap	STAMP
(i) Goals and objectives	Study 1 (Lee et al., 2017) To identify the contributing causes as well as interactions between them Sociotechnical systems theory	Study 2 (Kee et al., 2016) To analyse systematically the South Korea Ferry accident Sociotechnical systems theory
(ii) Theoretical background		Study 3 (H. Kim et al., 2016; T. Kim et al., 2016) To illustrate the utility of applying the STAMP to the maritime transportation domain Sociotechnical systems theory
(iii) Procedure	The accident analysis was carried out by the authors but there no evidence for additional review/validation	Study 4 (Kwon, 2016) To show the application of CAST to investigate the Sewol Ferry Accident Sociotechnical systems theory
(iv) Changes to standard Accimap/STAMP	No major changes	The accident analysis was carried out by the authors but there no evidence for additional review/validation No major changes
(v) System representation	One Accimap	One hierarchical safety control structure diagram No major changes
(vi) Levels of analysis	6 levels: outcomes; events & acts; technical & operational management; company management; regulatory bodies; government & legislation	Two separate control structure diagrams (one for capsizes and the other for rescue) 6 levels: crew and vessel; operation safety manager; ship-owning company; shipping association; regulatory bodies; government
(vii) Causal factors identified	28 factors	Two separate control structure diagrams (one for capsizes and the other for rescue) 10 levels: bridge; engine room; passenger cabin; vessel; ferry operation company; supporting companies; coast guard; shipping association; regulatory bodies; government 348 factors
(viii) Sources of data used	Safety investigation report by Korea Maritime Safety Tribunal (2014); news articles and columns in major South Korean newspapers and broadcast (2014)	37 factors Safety investigation report by Korea Maritime Safety Tribunal (2014); Progress on and Follow-up Measures for Sewol-ho Accident Recovery by Korea Ministry of Oceans and Fisheries (2014); three other published academic articles (2015)

transportation (Kee et al., 2016; Lee et al., 2017). Accimap analyses typically focus on failures across the following six organisational levels: government policy and budgeting; regulatory bodies and associations; local area government planning and budgeting (including company management) technical and operational management; physical processes and actor activities; and equipment and surroundings. Notably, Accimap is a generic approach and does not use taxonomies of failures across the different levels of analysis (Salmon et al., 2012; Waterson et al., 2017).

Systems Theory Accident Modelling and Process (STAMP) was formally introduced by Leveson (2004) and has subsequently been used in retrospective and prospective analyses in various domains, e.g., aerospace (Johnson and Holloway, 2003), led outdoor activities (Salmon et al., 2012) and marine transportation (T. Kim et al., 2016; Kwon, 2016). STAMP is constructed upon basic Systems Theory (Leveson, 2011) and focuses on inadequate control or enforcement of safety-related constraints on the system design, development and operation (Leveson, 2011). STAMP views systems as hierarchical structures with multiple control levels. Each level in the hierarchy imposes constraints on the activity of the level beneath it, the events leading to losses only occur when safety constraints were not successfully enforced or the constraints have been violated (Leveson, 2011). The potential for unsafe control may exist in the original design of the safety control structure and the controls may degrade over time, allowing the system to move to states of increasing risk. To support the identification of control failures, Leveson (2004) proposes a taxonomy of control failures, including inadequate control of actions; inadequate execution of control actions; and inadequate or missing feedback. Subsequent STAMP analyses have also included ‘mental model flaws’ in order to cater better for human control structures in the system as the origin of the method is in engineering

3.3. A framework for comparing the studies

In order to systematically compare the four studies, eight separate categories were used and adapted from Waterson et al. (2017), these were: (1) the goals and objectives of the study, e.g., providing a systemic account of the factors contributing to the accident; (2) the theoretical background of the study, e.g., sociotechnical systems theory; (3) the procedure described in building the Accimap and STAMP, e.g., review and validation; (4) changes to the standard Accimap and STAMP, i.e. major/minor deviations from the standard formats; (5) system representation: number of diagrams created; (6) levels of analysis, i.e., number and type of levels used; (7) total number of causal factors identified; and (8) the sources of data used.

The outputs of the four studies were evaluated qualitatively and quantitatively in terms of:

- The number of causal factors identified at each level and in total;
- The number of common factors between studies with the same methods (e.g., Lee et al., 2017: Accimap and Kee et al., 2016: STAMP) and different methods (e.g., Lee et al., 2017: Accimap and T. Kim et al., 2016: STAMP) at each level and in total;
- Overlap percentage of causal factors (the percentage of causal factors identified in a less detailed analysis overlap with causal factors identified in a more detailed analysis) For example, Study 1 and 2 identified 21 and 40 causal factors respectively with 8 common factors. It means 38% of causal factors identified in Study 1 overlap with factors identified in Study 2;
- Causal links between the factors;
- Safety recommendations and the parties at which the recommendations were directed.

The qualitative analysis was conducted by three authors, who have extensive experience in accident analysis method. They independently analysed the results from four studies. The causal factors and

recommendations were identified, compared and analysed. It was noticed that different analysts may use different language when referring to the same causal factors and recommendations and may differ with respect to whether these items are listed separately or grouped together. As a result, it was necessary for judgements to be made regarding whether or not differently-phrased and differently-formatted items refer to the same items. In this study, the judgements were made by three authors of this study. When the judgments were different, the factor was discussed until the consensus is achieved. For example, Study 1 identified “*Captain and some of the crew members left the ship without telling the other crew members or passengers to abandon the ship or release lifeboats*” as a causal factor, while Study 2 identified “*Stay put and first to abandon ship*” as a cause. Although the wording is clearly different, it seems apparent that both analyses are referring to the same causal factor and the three authors concluded that these factors were same.

4. Findings

Table 1 summarises the comparison of four studies based on the framework. In subsequent sections of this paper, we refer to the studies with reference to their number in Table 1 (e.g., Study 1 is Lee et al., 2017).

The four studies have the same overall goal which was to apply a systems approach to analyse the South Korea Ferry accident and identify the range of contributing factors. Studies 1 and 2 applied the Accimap to analyse the accident and identify the contributing factors as well as interactions between them. Studies 3 and 4 applied the STAMP method to the entire maritime transportation socio-technical system in order to provide a broader view of factors contributing to the accident. Socio-technical systems theory was adopted as a theoretical background by all studies. Study 2 uniquely involved two method experts who created an initial Accimap, which was internally validated by the remaining co-author and externally validated by a senior operation manager in a shipping company. Studies 1, 3 and 4 did not provide details of the validation procedure. The four studies followed standard methodologies for Accimap and STAMP analyses without any major changes. Study 1 used the Accimap framework developed by Rasmussen and Svedung (2000) for the analysis of the Zeebrugge Ferry accident, whereas Study 2 adopted the Accimap structure used by Svedung and Rasmussen (2002) and Branford (2011). Both study 3 and 4 used the STAMP (CAST, Leveson, 2004) standard nine steps.

In terms of use of system representations, Study 1 and 3 created one system representation, whereas Study 2 and 4 created two separate representations dividing the accident analysis into two parts of the capsizing of the ferry and the rescue operation. Consequently, Study 2 and 4 have a higher level of detail, identifying 49 and 348 contributing factors respectively. Study 1 and 3 identified 28 and 37 factors, respectively. Both Study 1 and 2 analysed the six levels as described by Rasmussen (1997), but re-labelled level six to ‘outcome’. Studies 3 and 4 analysed the levels of the hierarchical safety control structure including the system development and system operation of ships. Study 3 analysed six levels (e.g., Ministry of Oceans and Fisheries; Korean Coast Guard; Korean Register of Shipping; and industry association; Ship-owning Company; Master and Crew) whereas Study 4 used four additional levels (e.g., Loading Service Company (LOSC); Life Rafts Inspection Company (LIC); Lashing Service Company (LASC); Passengers). In terms of data sources, all the four studies were based on investigation reports by Korean governmental agencies (either Korean Maritime Safety Tribunal or Board of Audit and Inspection of Korea). Study 1 and 2 used additional information from major South Korean and foreign newspapers and news magazines.

4.1. Reliability assessment through inter-analyst agreement

As can be seen in Table 2, Study 1 identified factors at 6 levels while

Table 2
Comparison between outputs from the study 1 and 2 – causal factors.

Actors	Study 1	Study 2	Common Factors	Overlap percentage* (%)
Ministry of Oceans and Fisheries (MOF)	1	4	–	–
Korean Register of Shipping (KRS)	1	2	–	–
Korean Coast Guard (KCG)	5	4	1	20
Korean Shipping Association (KSA)	1	2	–	–
Shipping Company	7	6	3	43
Master and Crew (MC)	6	10	4	67
Ministry of Security and Public Administration (MSPA)	–	3	–	–
Central Disaster and Safety Countermeasures Headquarter (CDSCH)	–	2	–	–
Regional Rescue Centre (RRC)	–	2	–	–
Incheon Regional Maritime Affairs and Port Administration (IRMAPA)	–	1	–	–
Culture of Society (CS)	–	4	–	–
Total	21	40	8	38%

* How many percentage of causal factors identified in a less detailed analysis (Study 1) overlap with causal factors identified in a more detailed analysis (Study 2).

Study 2 identified factors at 11 levels. Both studies only have common factors at the levels of Korean Coast Guard, shipping company and master/crew, where the percentages of overlap are 20%, 43% and 67%, respectively. Overall, eight common factors (38% of overlap) were identified.

Table 3 shows that Study 4 presents a much higher level of detail and identified 348 causal factors, whereas Study 3 identified 37 factors. The overlap percentage of Study 3 is as high as 65%.

The analysis of the ways casual factors are described/located and casual links are connected shows a great degree of subjectivity. In terms of factor description, Study 2 describes “Over-drained ballast water” and “Over-loaded cargo” in two different boxes, whereas Study 1 put them in one box. Study 1 described a specific legislation factor, e.g., “in 2008, under the Lee administration, the maximum allowable age for a passenger ship went to form 20 years to 30 years” but Study 2 described more broadly various factors such as “Culture of Society”, “Pro-Business Policy” and “Low Profitability.”

There were multiple examples showing that the same factor was located at different levels. For example, “Over-drained ballast water” and “Lack of proper securing” are at the level of physical/actor events, process and conditions in Study 2, but are at the level of technical and operational management in Study 1. A different causal link was identified to the same factor by Study 1 and 2. For instance, the factor “lack of proper securing” was caused by “poor safety operating check” in Study 2, whereas Study 1 identified “Cheonghaejin purchased an 18-year-old Japanese ferry ship that was soon going out of commission” as cause to “lack of proper securing”.

4.2. Validity assessment - the degree of similarity between Accimap and STAMP

Table 4 shows the degree of similarity of causal factors identified using different methods (Study 2 and 3). Study 2 and 3 were selected considering the very similar number of causal factors identified in each study (40 and 37 factors respectively). Between two studies, a very

Table 3
Comparison between outputs from the study 3 and 4 – causal factors.

Actors	Study 3	Study 4	Common Factors	Overlap percentage* (%)
Ministry of Oceans and Fisheries (MOF)	3	10	3	100
Korean Register of Shipping (KRS)	1	13	1	100
Korean Coast Guard (KCG)	3	92	2	67
Korean Shipping Association (KSA)	3	32	3	100
Shipping Company	12	41	7	58
Master and Crew (MC)	15	120	8	53
Service Company	–	27	–	–
Passenger	–	13	–	–
Total	37	348	24	65%

* How many percentage of causal factors identified in a less detailed analysis (Study 3) overlap with causal factors identified in a more detailed analysis (Study 4).

limited number of common factors were identified with Korean Coast Guard, Shipping Company and Master and Crew, so the overlap percentage is as low as 8%.

4.3. Recommendations

Table 5 shows the main recommendations presented in the four studies. In Study 1 the recommendations focused on legislative and regulatory changes (e.g., the national government and local governments need to strengthen enforcement of rules and procedures). Study 2 focused on South Korean culture and their recommendations involve wide-scale changes to the way in which socio-political and economic pressures should be relaxed in order to promote safety. Study 3 recommended improvements to the hierarchical safety control structure that ensures safe development and operation of passenger ships in South Korea as a whole. Finally, Study 4 produced more comprehensive and detailed recommendations (44 in total), covering each level of the hierarchical safety control structure that ensures safe development and operation of passenger ships in South Korea.

5. Discussion

5.1. Summary of findings

The aim of this study was to systematically assess the validity and reliability of two accident methods (Accimap and STAMP). Compared with other studies with the similar aim (Hollnagel and Speziali, 2008; Underwood and Waterson, 2014; Waterson et al., 2017), this study carried out much more systematic comparison using both quantitative and qualitative analyses of outcomes and procedures of four different studies investigating the same accident. The results of the causal factor comparison indicate that the reliability (degree of overlap of causal factors identified from the same method, i.e. inter-analyst overlap) of STAMP (65%) is higher than Accimap (38%). Although these findings should be treated with caution given the limited sample size and

Table 4
Comparison between Accimap (Study 2) and STAMP (Study 3) outputs – causal factors.

Actors	Study 2	Study 3	Common Factors	Overlap percentage* (%)
Ministry of Oceans and Fisheries (MOF)	4	3	–	–
Korean Register of Shipping (KRS)	2	1	–	–
Korean Coast Guard (KCG)	4	3	1	33
Korean Shipping Association (KSA)	2	3	–	–
Shipping Company	6	12	1	8
Master and Crew (MC)	10	15	4	27
Ministry of Security and Public Administration (MSPA)	3	–	–	–
Central Disaster and Safety Countermeasures Headquarter (CDSCH)	2	–	–	–
Regional Rescue Centre (RRC)	2	–	–	–
Incheon Regional Maritime Affairs and Port Administration (IRMAPA)	1	–	–	–
Culture of Society (CS)	4	–	–	–
Total	40	37	3	8%

* How many percentages of causal factors identified in a less detailed analysis (Study 3) overlap with causal factors identified in a more detailed analysis (Study 2).

Table 5
Recommendation presented in four studies.

Actors	Study 1	Study 2	Study 3	Study 4
Whole system	1	5	2	–
Ministry of Oceans and Fisheries (MOF)	1	–	–	1
Incheon Regional Maritime Affairs and Port Administration (IRMAPA)	1	–	1	1
Korean Register of Shipping (KRS)	–	–	1	4
Korean Coast Guard (KCG)	–	–	–	9
Korean Shipping Association (KSA)	–	–	–	5
Shipping Company	–	1	3	9
Operation/cargo loading services	–	–	1	7
Master and Crew (MC)	1	–	1	–
Physical equipment	–	–	–	7
Total	4	5	9	44

particularly the very high number of causal factors identified in Study 4 relative to the three other studies, they suggest that the use of a more structured method like STAMP can help produce a more reliable accident analysis results. On the other hand, the validity (degree of overlap of causal factors identified from two different methods) shows as low as 8%. This finding makes us question whether the evaluations of the degree of similarity between results obtained from different methods alone provide meaningful insights into assessing the validity. Further research should be undertaken to holistically assess the validity using multiple approaches: evaluation of results against ‘gold standard’ (the external criteria against which the results of a method are assessed are in existence *prior* to the application of the method) or experts or in term of internal logic as proposed by Branford (2007).

The comparison of recommendations of four studies indicates that STAMP-based analyses may help produce a wider range of recommendations across multiple system levels while Accimap-based analyses tend to focus on whole system-related recommendations. Besides, two different sets of recommendations from two Accimap analyses showed different emphasis. For example, Study 1 highlighted the need for legislation improvement, whereas the majority of recommendations in Study 2 addressed wider societal issues. While each method has produced its own results, there was the element of being complementarity between Accimap and STAMP. There was a tendency that STAMP (Study 3 and 4) focused on providing an accurate hierarchical safety control structure, whilst Accimap (Study 1 and 2) can capture general whole system of ‘bigger picture’ factors (e.g., low profitability and small size of ferry business, culture of society, pro-business policy), which could be difficult to be captured using the STAMP’s hierarchical control structure. In addition, Accimap, unlike STAMP, links together factors within one level and between any level, which helps freely capture interactions and causality across different system levels. This feature helped to shed light in, for example, the downstream connections of the following factors across the marine transportation system in Study 2: ‘*inappropriate patronage practices*’ (level 1) lead to ‘*lack of independence in the*

position of marine operating inspector’ (level 2) lead to ‘*poor safety operation checks* (level 4) lead to ‘*lack of proper securing*’ (level 5) leads to ‘*cargo shift*’ (level 5) lead to ‘*capsized ferry*’ (level 6). By contrast, STAMP effectively captured the context in which decisions were made, as well as cognitive aspects that lead up to the accident (e.g., flawed mental models). This feature helped better understand why something went wrong. For example, Study 3 (STAMP) demonstrated that ‘*lack of experience and training of the Master and Crew*’ (context in which decision was made) led them to ‘*underestimate of the outcome of the sudden turn*’ (flawed mental model).

5.2. The role of analyst and other influences

Our findings highlight some similarities and some marked differences in terms of the different causal factors and recommendations which are described in the four studies. In many respects, it could be argued that this reflects the subjectivity involved in deriving four separate accounts of the Sewol accident. Seen from this point of view analysing the various factors that contributed to the disaster is less about looking for root causes or ‘*hunting for facts*’ (Dekker, 2011) and more akin to Turner’s (1995) description of the process of understanding accident causality as like ‘*shaking a kaleidoscope*’ in order to examine different causal configurations. Waterson et al. (2017) for example, argues that questions centred on the reliability and validity of the Accimap method may be missing the point. Rather than viewing Accimaps, and possibly STAMP, as objective tools for accident diagnosis, it may better to treat them as supporting idea-generation and brainstorming around causality, which likewise involves some degree of subjectivity and expertise (see also Sharples, 2017).

In line with other research (Hale et al., 1998; Svenson et al., 1999; Underwood and Waterson, 2013; Canham et al., 2018), the background and profile of the accident analyst may serve as one influence on how the ‘*kaleidoscope is shaken*’. Hollnagel and Speziali (2008) similarly argued that it is simply not possible to begin an investigation with a completely open mind, just as it is not possible passively to ‘*see*’ what is there. It was not possible to identify the background of the authors of the four studies, but it might be explained why the different focus was given by each study and why variation exists in the analysis outputs even when the same method was used. Therefore, careful consideration of analysts’ background and the environment they are in should be given when comparing different accident analysis methods and drawing conclusions. Study 4 (Kwon, 2016) for example, was a Master’s thesis and it is tempting to speculate that one of the reasons that study identified so many causal factors ($n = 348$) as compared to the other studies, was that it wasn’t subject to the space constraints of a typical scientific journal paper¹. Similarly, it might be argued that because the analysis was completed as part of a degree assignment it may have been written with a view to *demonstrating* knowledge of the STAMP method, rather than just alone *contributing* to our

¹ In total Kwon et al., (2016) runs to 174 pages of text.

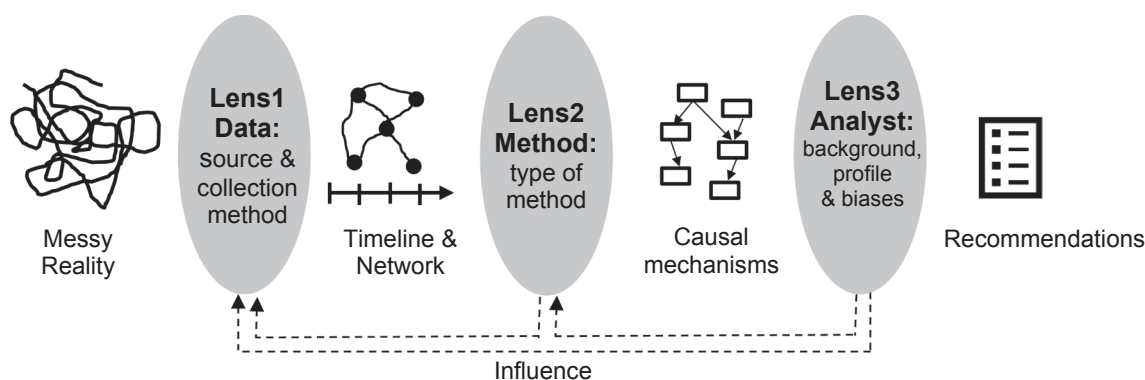


Fig. 1. Three 'lenses' which influence accident analysis.

knowledge of the Sewol accident.

It might also be argued that other preconceptions or biases (unconscious or conscious) on the part of the analyst(s) may explain the differences we found when the two methods were applied to the Sewol accident. The two STAMP studies, for example, set out partly to examine the feasibility and appropriateness of using the method in the maritime domain. By contrast, the two Accimap studies were less oriented around the method, but focused more on wider, systemic aspects of the accident and how these contrasted with prevalent ways of viewing accidents in terms of 'blame' (Study 2 – Lee et al., 2017) and the relationship between organisational learning and accountability (study 2 – Kee et al., 2017). Again, it is difficult to be precise about how preconceptions, theoretical preferences and other possible sources of bias may have crept into the four studies, but it is possible that a variation of what Lundberg et al. (2009) called the 'What-you-look-for-is-what you-find' principle may have been in operation such as the availability of different source data. A similar process may have applied to the case of various similarities and differences in terms of the recommendation for improvement generated by each of the studies (Lundberg et al., 2010). Fig. 1 summarises the discussion of this study by visually representing how multiple factors may shape the accident analysis process using the concept of three lenses consisting of i) data (data source & data collection method), ii) method (type of accident analysis method, use of systemic methods) and iii) analyst (background, profile and biases). Further work is needed to examine these and other influences on the use of systemic accident analysis methods.

6. Conclusions and future work

This study sets out to systematically assess the validity and reliability of two systemic accident analysis methods (Accimap and STAMP) by comparing both quantitative and qualitative outputs of four studies which analysed the same accident (South Korea Sewol Ferry accident). The study has shown that STAMP (68% overlap of causal factors) tends to be more reliable than Accimap (35% overlap). The study has also shown that STAMP and Accimap identify a very different set of causal factors (8% overlap). While each method has produced its own results, there was the element of being complementary to each other. Accimap captured soft factors like the culture of society and link them with factors in any other level, but STAMP focused on providing decision making and it context in an accurate hierarchical safety control structure. STAMP was particularly effective in producing a wider range of recommendations across multiple system levels while Accimap tends to focus on whole system-related recommendations. Considering the unique features of each method, the combined application, though might be practically challenging, can potentially help understand the entire mechanism and contributing factors of the accident when we investigate an accident. Further work needs to be done to assess the utility (e.g., training time, time-consuming, tools needed, input methods) and

ease of use/understanding of these methods as well as the reliability and validity.

Acknowledgment

The work described in this paper was carried out whilst the first author was an academic visitor in the 'Human Factors and Complex Systems Research Group' at Loughborough University during the period 2016–2017. We would like to thank Professor Nancy Leveson for sharing one of her MSc student's dissertations which was analysed in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssci.2018.12.002>.

References

- Branford, K., 2007. An Investigation into the Validity and Reliability of the Accimap Approach. PhD Thesis. Australian National University, Canberra.
- Branford, K., 2011. Seeing the big picture of mishaps: applying the Accimap approach to analyse system accidents. *Aviation Psychol. Appl. Human Factors* 1 (1), 31–37.
- Canham, A., Jun, G.T., Waterson, P.E., Khalid, S., 2018. Integrating systemic accident analysis into patient safety investigation practices. *Appl. Ergonom.* 72, 1–9.
- Dekker, S.W.A., 2011. Drift into Failure: From Hunting Broken Components to understanding complex systems. Ashgate, Farnham.
- Düzgün, H.S., Leveson, N., 2018. Analysis of soma mine disaster using causal analysis based on systems theory (CAST). *Safety Sci.* 110, 37–57.
- Goode, N., Salmon, P.M., Taylor, N.Z., Lenné, N.G., Finch, C.F., 2017. Developing a contributing factor classification scheme for Rasmussen's Accimap: Reliability and validity evaluation. *Safety Sci.* 64, 14–26.
- Gordon, R., Flin, R., Mearns, K., 2005. Designing and evaluating a human factors investigation tool (HFIT) for accident analysis. *Safety Sci.* 43, 147–171.
- Hale, A., Heming, B., Smit, K., Rodenburg, F., van Leeuwen, N., 1998. Evaluating safety in the management of maintenance activities in the chemical process industry. *Safety Sci.* 28, 21–44.
- Hollnagel, E., 2004. Barriers and Accident Prevention. Ashgate, Farnham.
- Hollnagel, E., Speziali, J., 2008. Study on Developments in Accident Investigation Methods: A Survey of the "State-of-the-Art". *SKI Report 2008:50*.
- Isaac, A., Lyons, M., Bove, T., Van Damme, D., 2003. Validation of the Human Error in ATM (HERA-JANUS) Technique. Eurocontrol, Brussels.
- Jacinto, C., Aspinwal, E., 2004. WAIT (Part III) - Preliminary validation studies. *Safety Sci. Monitor* 8, 19–29.
- Johnson, C.W., de Almeida, I.M., 2008. An investigation into the loss of the Brazilian space programme's launch vehicle VLS-1 V03. *Safety Sci.* 46 (1), 38–53.
- Johnson, C.W., Holloway, C.M., 2003. The ESA/NASA SOHO mission interruption: using the STAMP accident analysis technique for a software related 'mishap'. *Software – Pract. Exp.* 33, 1177–1198.
- Jun, G. [Gyuchan Thomas Jun], 2015, July 12. Two Contrasting Views of the South Korea Ferry Accident. Retrieved from <https://youtu.be/iZwbn8Y1Ywc>.
- Kassarjian, H.H., 1977. Content analysis in consumer research. *J. Consumer Res.* 4, 8–18.
- Katsakiori, P., Sakellaropoulos, G., Manatakis, E., 2009. Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. *Safety Sci.* 47, 1007–1025.
- Kee, D., Jun, G.T., Waterson, P.E., Haslam, R., 2017. A systemic analysis of South Korea Sewol Ferry accident - striking a balance between learning and accountability. *Appl. Ergonom.* 59 (B), 504–516.
- Kim, H., Haugen, S., Utne, I.B., 2016a. Assessment of accident theories for major

- accidents focusing on the MV SEWOL disaster: similarities, differences, and discussion for a combined approach. *Safety Sci.* 82, 410–420.
- Kim, T., Nazir, S., Øvergård, K.I., 2016b. A STAMP-based causal analysis of the Korean Sewol ferry accident. *Safety Sci.* 83, 93–101.
- Kirwan, B., 1992. Human error identification in human reliability assessment: Part 2: detailed comparison of techniques. *Appl. Ergonom.* 23 (6), 371–381.
- Kwon, Y., 2016. System Theoretic Safety Analysis of the Sewol-Ho Ferry Accident in South Korea. MSc Thesis. Engineering Systems Division, Massachusetts Institute of Technology, USA. Available at: <http://sunnyday.mit.edu/papers/Kwon-Thesis.pdf> (last accessed April 23rd, 2018).
- Lee, S., Moh, Y.B., Tabibzadeh, M., Meshkati, N., 2017. Applying the Accimap methodology to investigate the tragic Sewol Ferry accident in South Korea. *Appl. Ergonom.* 59 (B), 517–525.
- Leveson, N., 2004. A new accident model for engineering safer systems. *Safety Sci.* 42, 237–270.
- Leveson, N.G., 2011. *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press, Cambridge, MA.
- Lundberg, J., Rollenhagen, C., Hollnagel, E., 2009. What-you-look-for-is-what-you-find – the consequences of underlying accident models in eight investigation manuals. *Safety Sci.* 47, 1297–1311.
- Lundberg, J., Rollenhagen, C., Hollnagel, E., 2010. What you find is not always what you fix – how other aspects of accidents decide recommendations for remedial actions. *Acc. Anal. Prevent.* 42, 2132–2139.
- Maxwell, J.A., 1992. Understanding validity in qualitative research. *Harvard Educ. Rev.* 62 (3), 279–300.
- Militello, L.G., Hutton, J.B., 1998. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics* 41 (11), 1618–1641.
- Nayak, R., Waterson, P., 2016. 'When Food Kills': a socio-technical systems analysis of the UK Pennington 1996 and 2005 E. coli O157 Outbreak reports. *Safety Sci.* 86, 36–47.
- Pounds, J., Isaac, A., 2003. Validation of the JANUS Technique: Causal Factors of Human Error in Operational Errors. Department of Transportation, Washington, U.S.
- Rasmussen, J., 1997. Risk management in a dynamic society: a modeling problem. *Safety Sci.* 27 (2), 183–213.
- Rasmussen, J., Svedung, I., 2000. *Proactive Risk Management in a Dynamic Society*, first ed. Raddningsverket, Risk and Environmental Department. Swedish Rescue Services Agency, Karlstad, Sweden.
- Ryan, B., 2015. Incident reporting and analysis. In: Wilson, J.R., Sharples, S. (Eds.), *Evaluation of Human Work* (4th Edition). CRC Press, Boca Raton.
- Salmon, P., Cornelissen, M., Trotter, M.J., 2012. Systems based analysis methods: a comparison of Accimap, HFACS and STAMP. *Safety Sci.* 50, 1158–1170.
- Schmid, D., Vollrath, M., Stanton, N.A., 2018. The System Theoretic Accident Modelling and Process (STAMP) of medical pilot knock-out events: Pilot incapacitation and homicide-suicide. *Safety Sci.* 110, 58–71.
- Sharples, S., 2017. Commentary: analysis, investigation and judgement: the post-hoc application of human factors analyses to incidents. *Appl. Ergonom.* 59, 526–527.
- Stanton, N.A., Young, M.S., 1999. What price ergonomics? *Nature* 399, 197–198.
- Stanton, N.A., Young, M.S., 2003. Giving ergonomics away? The application of ergonomics methods by novices. *Appl. Ergonom.* 34, 479–490.
- Stevens, N.J., Salmon, P.M., 2016. Sand, surf and sideways: a systems analysis of beaches as complex roadway environments. *Safety Sci.* 85, 152–162.
- Svedung, I., Rasmussen, J., 2002. Graphic representation of accident scenarios: mapping system structure and the causation of accidents. *Safety Sci.* 40, 397–417.
- Svenson, O., Lekberg, A., Jonhansson, A.E.L., 1999. On perspective, expertise and differences in accident analyses: arguments for a multidisciplinary integrated approach. *Ergonomics* 42 (11), 1561–1571.
- Turner, B., 1995. A personal trajectory through organization studies. *Res. Sociol. Organ.* 13, 275–301.
- Underwood, P., Waterson, P.E., 2013. Systemic accident analysis: examining the gap between research and practice. *Acc. Anal. Prevent.* 55, 154–164.
- Underwood, P., Waterson, P.E., 2014. Systems thinking, the Swiss Cheese Model and accident analysis: a comparative systemic analysis of the Grayrigg train derailment using the ATSB, AcciMap and STAMP models. *Acc. Anal. Prevent.* 68, 75–94.
- Valdez Banda, O.A., Goerlandt, F., 2018. A STAMP-based approach for designing maritime safety management systems. *Safety Sci.* 109, 109–129.
- Waterson, P.E., 2009. A systems ergonomics analysis of the Maidstone and Tunbridge Wells infection outbreaks. *Ergonomics* 52, 1196–1205.
- Waterson, P.E., Robertson, M.M., Cooke, N.J., Militello, L., Roth, E., Stanton, N.A., 2015. Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety. *Ergonomics* 58, 650–658.
- Waterson, P.E., Jenkins, D.P., Salmon, P.M., Underwood, P., 2017. 'Remixing Rasmussen': the evolution of Accimaps within systemic analysis. *Applied Ergonomics* 59 (Part B), 483–503.
- Zhou, W., Zhao, T., Liu, W., Tang, J., 2018. Tower crane safety on construction sites: a complex sociotechnical system perspective. *Safety Sci.* 109, 95–108.