

Applying a longitudinal tracer methodology to evaluate complex interventions in complex settings

Paul Fuller^a, Ray Randall^b, Andrew Dainty^a, Roger Haslam^c & Alistair Gibb^{a*}

^a School of Architecture, Building & Civil Engineering, Loughborough University, Loughborough, UK. ^b School of Business and Economics, Loughborough University, Loughborough, UK. ^c Loughborough Design School, Loughborough University, Loughborough, UK.

*corresponding author - Alistair Gibb a.g.gibb@lboro.ac.uk Loughborough University

Applying a longitudinal tracer methodology to evaluate complex interventions in complex settings

Long-running multi-faceted intervention studies are particularly problematic in large complex organizations where traditional methods prove too resource intensive and can yield inaccurate and incomplete findings. This paper describes the first use of, longitudinal tracer methodology (LTM), a realist approach to evaluation, to examine the links between multiple complex intervention activities (processes) and their outcomes on a construction megaproject. LTM is especially useful when the researcher has little control over intervention delivery but has evidence drawn from multiple sources to evaluate the intervention activities effects over time. This methodology has rarely been deployed in complex organisational settings and not on a construction megaproject. This paper presents a case study of its use over a period of three years, on 24 construction sites forming London's Thames Tideway Tunnel (Tideway) megaproject. The study examines the 'transformational' power of occupational safety and health (OSH) interventions across the multiple organisations and supply chains in the megaproject. The study shows how the method can be adapted in-flight to accommodate shifting lines of inquiry as the intervention activities progress and change. This feature along with its resource efficient operation, make it any attractive option where interventions are likely to have differential effects across multiple sites of enactment.

Keywords: Evaluation; longitudinal; tracer methodology; occupational safety and health; realist; organisational change.

Introduction

Organisational change research and practice is characterised by long implementation periods, the involvement of multiple stakeholders in intervention design and implementation and fluctuating influential organisational contexts (Johns, 2006; 2018; Nielsen & Miraglia, 2017; Rossi, Lipsey, & Freeman, 2003). Unfortunately, as Tsoukas and Chia (2002) note, the evaluation of complex organisational change interventions has been “dominated by assumptions providing stability, routine and order.” (p.567). To better reflect the actuality of change interventions they proposed treating change as dynamic and never complete, arguing for methods that are robust to “the reweaving of actors’ webs of beliefs and habits of action to accommodate new experiences obtained through interactions.” (p.567). This implies that change interventions will frequently:

- (1) unfold and change over time during their adaption from plans into practice (Mintzberg, 1987);
- (2) be changed by the people and context in which they are deployed (von Thiele Schwarz, Lundmar, & Hasson, 2016; Weiner, 2009) and;
- (3) occur within shifting organizational and operational contexts (Johns, 2006; Nielsen & Randall, 2013).

Interventions in organisations are often the result of the implementation of a strategy developed by those in leadership positions and other with expertise to change the way the organisation operates in practice. The Strategy as Practice (SaP) literature echoes the organisational change proposition that interventions often do not go according to plan. As a result, there are a number of potential, sometimes unexpected, outcomes that may result from implementing a strategy (Mirabeau et al., 2017,

Jarzabkowski et al., 2016). In Mintzberg's (1987) typology there are five basic types of strategy: intended; deliberate; unrealized; emergent and; realized. He stated that several of these may occur concurrently and consecutively in complex interventions making evaluation difficult. A sixth type, ephemeral, is an emergent strategy that disappears and further underscores the complexity of evaluation (Mirabeau & Maguire, 2014). There have also been calls for macro perspectives (impacts external to an organisation) to be examined during evaluation (Seidl & Whittington, 2014, Jarzabkowski, & Spee, 2009). In this article we apply the notion of SaP to a LTM case study of a transformational OSH strategy on London's Thames Tideway Tunnel construction megaproject, generally just called Tideway. The project is highly complex, has multiple delivery partners and has macro impacts on wider society. The intervention implemented is designed to influence practices both within the project and across wider industry. The three-year LTM research project, the first of its kind world-wide, was funded by the Institution of Occupational Safety and Health (IOSH).

The evaluation of such projects is challenging. Traditional evaluation methods often rely on planned and controlled delivery of pre-determined interventions (Cook & Campbell, 1979). Only stable intended and deliberate interventions are readily tracked through such approaches to evaluation, for example quasi-experiments. Such approaches to evaluation are usually resource intensive and sometimes disruptive. This limits their suitability for use by practitioners and researchers working with large multi-faceted interventions, over extended periods of time in turbulent and complex organisational contexts (Bryman, 2012; Pawson & Tilley, 1997; Randall, Griffiths, & Cox, 2005).

One way of overcoming some of these difficulties is to review and adapt the evaluation methods and design over time so that the changes can be quickly and

accurately observed and recorded (Rossi, Lipsey, & Freeman, 2003). This realist approach to evaluation involves the collection and use of information about shifts in intervention plans, activities and contexts to shape the evaluation of intervention effectiveness (Pawson, 2013; Pawson & Tilley, 1997). Building on the concepts of SaP we describe a practical approach to realist evaluation employing LTM. We show how it can be used to identify appropriate evaluation criteria and to uncover why, for whom, against which criteria, and under what circumstances, unfolding interventions work (or fail). This type of information is vital if interventions are to be transferred to new settings (Pawson & Manzano-Santaella, 2012)).

OSH strategy on Tideway is a complex intervention with transformational intended outcomes being implemented by multiple stakeholders, over a significant period of time, within a complex and rapidly changing organisational context. Evaluating the effectiveness of this type of intervention is problematic but vital (Nielsen, 2017; van der Molen et al., 2018). There remains significant debate about the effectiveness of these complex interventions in industry and, in particular, construction (Hale, 2014; Harrison & Dawson, 2016; Nielsen et al., 2010; Sherrat, 2018; Waterman 2007; Whysall, Haslam, & Haslam, 2006). LTM might offer a practical and yet rigorous enough approach to address these important challenges.

The following sections of this paper describe realist evaluation and LTM. We then describe the application of the approach to OSH interventions in a construction megaproject case study where SaP needs to be examined. This is followed by a discussion on the challenges presented by the context of the project and how they were overcome. We conclude with a summary of the key learning points for researchers and practitioners.

Realist evaluation: An overview

There are very significant challenges when evaluating complex interventions in work organizations. Quasi-experiments are often cited as the most informative approach when randomization of participants to intervention and control groups cannot be achieved (Cook & Campbell, 1979). The explanatory power of quasi-experiments is contingent upon the delivery of carefully managed patterns of exposure to interventions with few components (Pawson, 2013). However, researchers in organizations are usually faced with the challenge of evaluating evolving and more complex interventions taking place in environments that themselves are evolving and complex. Typically, OSH interventions have multiple and linked working components, multiple stakeholders (designing, delivering and receiving the intervention) and have multiple pathways to multiple outcomes (Hale, 2014). Workers' exposure to and experiences of interventions are susceptible to the influence of a myriad of factors outside of the researchers' control. Evaluating complex organizational change has proved difficult for researchers because of the weakening of quasi-experimental designs that results from intervention and contextual complexity (Griffiths, 1999).

Realist evaluation is designed to be robust and informative in circumstances that prevent certainty of evaluation (Pawson, 2013). It is a flexible, thorough and continuous form of enquiry that utilises information about unfolding complexities to develop evidence-based explanations for intervention effects. Intervention and contextual complexity are treated as valuable data that can be used to shape evaluation and not as sources of imperfections in the research design or confounding variables (Griffiths, 1999; Nielsen & Randall, 2013; Rossi, Lipsey, & Freeman, 2003). Crucially it releases the evaluation methodology from the constraints of controlled and carefully managed intervention exposure. This is especially useful when control over

intervention activities and contexts is limited by factors such as: long intervention periods; intervention design and delivery occurring in open (rather than closed) organisational systems; the risks and costs associated with the controlled use of pre-determined intervention and control groups; and the potential for research activities to cause disruption to participants' on-going work activities (Randall, Griffiths, & Cox, 2005; von Thiele Schwarz, Lundmar, & Hasson, 2016). In OSH research, realist evaluation has been used to measure diverse and unpredictable stakeholder perceptions of, and attitudes and behavioural responses to, many change activities (Nielsen & Randall, 2013; Nytrø, Saksvik, Mikkelsen, Bohle, & Quinlan, 2000; Randall, Nielsen, & Tvedt, 2009)). There is now ample evidence that such information can be used to develop robust explanations for intervention successes and failures (Fridrich, Jenny, & Bauer, 2016, Hasson, von Thiele Schwarz, Nielsen, & Tafvelin, 2016).

Using a longitudinal approach allows the time order of events to be observed which can allow causal inferences to be made more reliably compared to cross-sectional methods. (Bryman, 2012). Realist evaluation offsets lack of researcher control through careful examination of the unfolding heterogeneity and complexity of the intervention experiences of those it impacts (Pawson, 2013). Causal inferences are made more robust by collecting data frequently and from a variety of reliable and meaningful sources and by using multiple methods. These longitudinal mixed methods approaches afford rich insights into SaP by illuminating the ways in which planned intervention strategies evolve over time into practical activities. These differences between intervention plans and practical activities are shaped by stakeholders' knowledge, decisions made by professional practitioners and changes in the environments in which the activities take place (Jarzabkowski & Spee, 2009). As

such, comprehensive realist evaluation goes beyond an assessment of the fidelity of the intervention (the extent to which the enacted intervention matches the intervention plan) to encompass the collection of data about when, why and how intervention activities went as planned or were modified or omitted. Data to support this evaluation can come from researchers' observations, documentary evidence of intervention activities and self-report data from those involved in the design and delivery of the intervention. Realist evaluation methods can capture recipients' knowledge, perceptions and appraisals of health-related intervention activities (Havermans, Schelvis, Boot, Brouwers, Anema, & Van der Beek, 2016). These are common mediators in program theory that are often overlooked in outcome-focused evaluation of occupational health interventions (Nielsen & Miraglia, 2017).

Realist evaluation requires the specification of program theory that connects intervention-related events and intervention outcomes. This involves the development of models of the linkages, or causal pathways, between intervention activities, intervention resources, intervention contexts and both short-term and long-term intervention outcomes (Rogers, 2008). These pathways often remain opaque 'black boxes' in outcome-focused evaluation (Nielsen & Randall, 2013). The application of the principles of SaP requires the identification and testing of modified and new pathways that might have emerged from the intervention activities (e.g. if an intervention was poorly or inconsistently delivered, modified or appeared to have unintended consequences). This can be achieved through rigorous data collection and analysis throughout the intervention process, from initiation through to delivery and maintenance / adjustment of the intervention and eventual end-point (Rossi, Lipsey, & Freeman, 2003). Emergent comparison groups then become evident through

variations in intervention activities, resources and contexts (Randall, Griffiths, & Cox 2005).

In summary, realist evaluation moves the focus away from the outcome-focused question ‘does it work?’ to a deep and broad investigation of the change mechanisms driving and shaping intervention effects. Specification and re-specification of program theories as interventions unfold captures complexity and provides detailed insights into the reasons for their effects. These data can also be used to guide adjustments to interventions to address problems that may be undermining its outcomes (von Thiele Schwarz, Lundmar, & Hasson, 2016).

Realist evaluation using LTM

Realist evaluation methods have the potential to show how complex organisational change phenomena unfold over time to impact in different ways for different stakeholder groups (Chau and Witcher, 2005, Combey, 1980; Hornby & Symon, 1994; Woodward, 1970;). There remain, however, relatively few such studies which reveal their effectiveness. Jarzabkowski and Spee (2009) argue that this may be because of “the type of time investment necessary to collect and analyse practice-based data sets which typically are longitudinal, rich and qualitative” (p.91). LTM is an example of realist evaluation that can be both a rigorous and resource efficient way (Combey, 1980). The following sections provide an overview of LTM and examine its application to the evaluation of a complex OSH intervention taking place within a construction megaproject. Interventions within megaprojects are challenging to evaluate. These projects are very large, highly complex costing over US\$1 billion with multi-year development and build lifecycles, involving multiple stakeholders (private and public), have transformational goals, impacting millions of people (Flyvberg, Bruzelius, & Rothengatter, 2003).

An initial program theory is a “causal modal linking programme inputs and activities to a chain of intended or observed outcomes, and then using this model to guide the evaluation” (Rogers, 2008, p. 30). This is the basis for the design and execution of LTM. In OSH interventions program theories often include changes in behaviour and in the contexts and events that connect changes in behaviour to OSH related outcomes. These might include the transfer of training into the workplace, a good prevailing safety climate, adequate access to good quality and appropriate personal protective equipment (PPE) and so on. Program theory is used as the basis of a specification for measurement and analysis.

LTM, although rarely used, is not new. The approach was initially developed by Woodward (1970), who used it as a method to explore the effects of complex computer-based managerial control systems. Manufacturing processes across three case study sites were selected as ‘tracers’ of the effects of implementing the control systems. Employee behaviour and worker interactions were observed over time to gather data on the unfolding effects of the control systems on the planning of work activities, decision-making and the way manufacturing tasks were carried out. The LTM, in this example, involved intense data collection focusing on small elements of activities that were identified by stakeholders as being susceptible to the effects of the intervention. Importantly, this focus shifted between different sub-systems and activities as the intervention evolved and changed. This flexible approach supported the development of broad and deep understanding of the dynamic, and sometimes unexpected, ways in which the control systems were evident and exert their effects.

In terms of the advantages offered by tracer studies, Hornby & Symon (1994) argued that:

“...tracer studies uniquely allow the investigator to study complex processes over time, observing the interconnectedness of episodes and issues and examining participants’ ideas, motives, meanings and perceptions as opposed to simply measuring their attitudes to outcomes or the relationships between members of temporary groupings.” (p.184).

LTM requires on-going dialogue with multiple stakeholders alongside the collection and analysis of multiple sources of data about intervention process and outcome. A variety of research techniques are employed e.g. interviews, direct observation, document analysis, questionnaires. ‘Tags’ are applied to items that provide evidence of the implementation of the intervention (either as intended in the program theory or in a modified way based on information from stakeholders), its effects and the interplay between intervention activities and the wider organisational context. Examples of items that can be ‘tagged’ include organisational policies, minutes of meetings, researcher observations or records of the behaviour of those interacting with the intervention. These tagged items are then followed (traced) to reveal the unfolding intervention processes and outcomes over time.

LTM approaches can reduce costs and disruption because tags are applied to activities that are an integral part of the intervention and its context (as opposed to a requirement of a research methodology). As complex interventions evolve, following the tags and tracers over time helps to identify developing, testable, causal connections between intervention activities, intervention contexts and intervention outcomes. Well-selected tags and tracers provide rich information about the social interactions of the actors involved. This means that tight control over the delivery of the intervention is not needed to produce usable evaluation data. LTM retains two main advantages of longitudinal research: inaccurate program theories are identifiable

as change processes are observed over time; and the time priority of observed changes provides insights into causal effects. As Miller and Friesen (1982) argued:

“A broad knowledge of the nature of organizations, its environments, the personalities of its managers, etc. allows researchers to make inferences about why things happened. They can see the specific processes that lead up to critical events and can much more readily distinguish between cause and coincidence.” (p.1014)

Regular data collection is a particularly important element of LTM. It avoids the pitfalls associated with the infrequent collection of cross-sectional snap-shots with findings being extrapolated backwards and forwards over time and outside of the timeframe the data was collected (Yin, 2003). This approach also tends to be remote from intervention events and does not provide the rich insights into the context of changes as they occur. LTM overcomes these limitations as it does not rely on a combination of the potentially inaccurate or biased recollections of research subjects or on data derived from lagging indicators.

The application of LTM in a construction megaproject

Most of the research into OSH interventions in complex multi-site construction projects with networked supply chains has been cross-sectional (Hale, 2014). In contrast, LTM captures the ways in which OSH policies and practices intersect and intertwine with other activities taking place within a complex and fluctuating organisational context.

These projects typically have a temporary multi-organisational (TMO) structure leading to complex, unstable and evolving contractual arrangements and changing organisational structures (Baccarini, 1996; Davies & Mackenzie, 2014). This makes LTM particularly suitable as it is essential to observe any changes

affecting interventions as they occur over the project lifecycle. Construction projects in general are becoming more complex with multiple joint ventures as elements of TMO'S now common. The Tideway project will construct a 25 km, 7.2m diameter interception, storage and transfer tunnel running up to 65 metres below the river Thames with a capacity of 1.6 million cubic metres and at a cost of around £4.2bn. Starting in west London, the main tunnel generally follows the route of the river to Limehouse, where it then continues north- east to the Abbey Mills Pumping Station near Stratford. There it will be connected to the Lee Tunnel, which will transfer the sewage to Beckton Sewage Treatment Works.

The work is divided up into three sections, west, central and east, each of which will be constructed by a different joint venture team, led by British, French and Spanish parent company organisations. A separate organisation is responsible for the overall operational sewage system control integration. There is a management company, acting on behalf of the ultimate client, which is supervising construction.

The Tideway client has an aspiration to deliver a transformational approach to OSH which enables the sharing of best practice to deliver health and safety performance better than any else currently experienced in construction. This project provides a unique opportunity to study the impact of OSH leadership, policy and practice over an extended period during its delivery over the first three years of an eight-year project.

The intervention being studied began in 2016 (Fuller et al., 2017). LTM follows the processes and people that emerge as relevant to shaping OSH outcomes. This is in contrast to other evaluation approaches that at the outset specify and fix the factors (i.e. the program theory) thought to be related to OSH outcomes.

- Insert Figure 1 about here-

As Figure 1 illustrates, LTM allows a variety of research avenues to be opened and closed as the relevance of each is tested. Thus, examining the effect of specific interventions as ongoing and changing activities. Data analysis informs iterative examination of emergent issues by triggering modifications to the data collection strategy as the intervention unfolds (Chau & Witcher, 2005). The choice as to which issues are to be examined allows data collection to be focussed in specific areas of interest, making the dataset more manageable and evaluation more resource efficient. It also enables a better understanding of the wider picture through looking at small elements of the organisation rather than all elements at one point in time.

In the case study the data has been drawn from a variety of techniques to carry out LTM. Some are collected through bespoke techniques and measures (e.g. questionnaires, interviews and focus groups) and some are observational data (notes from researchers). An important advantage of LTM is that much comes from information routinely collected by the organisation (including employee absence data, organisational charts, project delivery timetables, records of personnel changes etc.).

Examining Strategy as Practice using LTM

LTM can be used to identify how and why strategy evolves into practice (Jarzabkowski 2004; Pettigrew, 1990, 1992; Whittington 2006). The examination of how strategy becomes practice (Pettigrew, 1990) allows the various OSH policy strands to be traced from their development through to adoption as practice.

Jarzaborwoski (2004) proposed that strategy is traditionally seen as something that an organisation has. In contrast, SaP conceptualises strategy as something people (in organisations) do (in other words, people influence the shape of the plan as it moves through implementation). More specifically, strategy practitioners make, shape and

execute strategies. The implication is that what people do in practice, strategy praxis, can differ significantly from the articulated strategy.

Mintzberg (1987) pointed out that that the commonly accepted meaning of strategy was that it was a plan made in advance involving “some sort of *consciously intended* course of action, a guideline (or set of guidelines) to deal with a situation.” (p.11). However, he also argued that strategy is “a pattern in a stream of actions.” (p.12) which takes into account realisation of the strategy in terms of the behaviour it drives. The plan and the strategy can be independent of each other, with the plan being the intended strategy and the pattern being the realized strategy. Figure 2 shows how deliberate strategies are the result of intended strategies, but that other emergent strategies can develop without pre-existing intentions. Other intended strategies may not translate into action (unrealized strategies).

-Insert Figure 2 about here-

Tsoukas & Chia (2002) argue that change programs evolve through a number of processes that need to be documented and analysed in order to understand fully the effects of change. They proposed that change program outcomes are difficult to predict as they are changed by the actors involved during their implementation.

An example intervention from within the Tideway megaproject covering both branding and OSH illustrates these different types of strategies and how they evolve over time. The intended strategy was described in the Works Instructions (WI) for PPE. This specified the need for the use of industry standard PPE e.g. high visibility orange safety clothing, safety work boots, protective gloves and so on. This program theory was then influenced by emergent strategies and patterns of actions that reflected stakeholders’ desire to be more ‘transformational’ and develop project specific red and teal clothing and high specification safety boots (with snowboard-

boot style lacing). This made relevant procurement process and issues around the clothing itself, so that current practice on site is transitional (mix of red and teal, orange). Figure 3 shows these changes incorporated into the Mintzberg model.

-Insert Figure 3 about here-

LTM revealed that the plan was modified. The strategy as practice is that the red and teal high visibility clothing, and high specification safety boots will only be mandatory for a specified sub-group of staff working on a specific part of the project and not the entire project staff as was originally planned. Mintzberg's model (1987) suggests that there could be further changes to the implementation of the PPE strategy elements. LTM is being used to collect data on these changes in real time. Quasi-experimental evaluation would support only a comparison between the final outcome and the original expectations as they appeared in the strategy-as-planned (the W I documentation that was a tag in this intervention). This fixed approach would provide limited information about what happened in the intervening period. At intervention evaluation, the main sources of data would be recollections of staff which can be reliable and valid. However, in construction projects attrition is high as many workers will have left the project or changed roles. This could lead to incorrect findings and conclusions being drawn (see Figure 4). The items in the grey box are potential items to be 'tagged' and 'traced' or followed during the study.

-Insert Figure 4 about here-

In terms of data collection, the project contains multiple activities on twenty-four sites, being delivered and received by multiple stakeholders. This means that it is not resource efficient to collect and analyse all available data. LTM follows initial lines of enquiry guided by the program theory, using information about the intended strategy, continuing strategies and patterns of practice. Knowledge of the

organisational context and information about deviations from the intended strategy (unrealised strategies) were used to guide scanning for emergent strategies resulting in unplanned but observable patterns.

Data collected at the strategy development and implementation phase includes information about how and why the interventions were chosen and developed. The program theory (intended strategy) upon which the initial LTM was based involved delivery of a sequence of packages of specific OSH interventions. This means that the reach of the intervention activities into the target population (continuing strategy and patterns of practice) can be readily monitored. Documents such as minutes of meetings, workshop material, communications and project reviews indicate when and where the intervention was delivered according to the intended strategy. Such data can then be used to identify emergent comparison groups (those the intervention had not yet reached) that enable outcome evaluation that reflects the continuing strategy (Randall, Griffiths, & Cox, 2005).

Alongside specifying and discovering change mechanisms, LTM supports an analysis of the interactions and transactions between the intervention and the contexts within which it takes place. This identifies reasons for the development of unrealised and emergent strategies. Many of the interventions studied were designed to have an impact on the day-to-day activities of individual workers. Those individuals each have their own work demands and routines (their individual work context); each individual works within a team context; and each team context exists within an organisational context; in a megaproject each organisation exists within the context of the megaproject i.e. a TMO. Large-scale interventions are often also impacted by, and indeed may influence, economic and political contexts outside of the organisation. These multiple contexts can have relatively stable background features such as the

organisational culture (the omnibus context) and more features more proximal to the intervention (the discrete context) that may fluctuate more frequently (Johns, 2006; 2018).

LTM utilises contextual data in two ways. First, data are collected about the impact of context on the initiation, design, delivery and maintenance of the intervention. Second, an assessment is made of the direct impact of these contexts on the outcomes specified in the program theory. Positive (facilitating) and negative (hindering) contextual factors can be identified (Nielsen & Randall, 2013). For example, in Tideway, changes in leadership or staff turnover (the discrete, team context) can significantly disrupt access to well-designed and well-resourced interventions. Decisions to accelerate project activities (the omnibus context) can mean that modifications to interventions are required.

Mintzberg's theory indicates that tracers of the intervention in a large and complex project will reveal multiple intervention mechanisms, or multiple context-mediators-outcome (CMO) pathways (see Pawson & Tilley, 1997). In other words, the contexts can determine the extent to which the elements of a program theory are delivered and received, and their connections to intervention outcomes. By capturing first hand contextual information, LTM facilitates an assessment of the likely suitability of the intervention for use in new contexts and yields information about how interventions might need to be adjusted to improve the chances of success.

Data collection and analysis in LTM

The long timeline of the Tideway project provides opportunities to both open new avenues of enquiry that are pertinent to the enactment of OSH policy across a complex project, and to close down areas which offer little by way substantive new insights into the OSH environment. The following research questions were used to inform the

collection of data using the LTM approach (see Fuller et al., 2017):

- How does OSH policy (intended strategy) evolve into continuing strategy and then into practice?
- How does OSH policy propagate through complex organisations associated with this type of large-scale construction project?
- How have the specific OSH interventions been implemented and managed and how effective have these been? This question focuses on examining the links between the management of the intervention process and its outcomes.
- Which findings will be of most relevance and most use to industry practitioners and the wider research community?
- How do people think and behave when experiencing the complexity of change in megaprojects?

For the qualitative data (e.g. interviews and observations) NVivo was used to develop a coding framework around topics and sub-topics. The topics for investigation were identified using the intended strategy, a timeline of key events at Tideway (developed using a specialist software package, Aeon Timeline) and models of intervention process evaluation (PE) from the relevant OSH literature (e.g. Nielsen & Randall, 2013).

The effectiveness of interventions in delivering intended outcomes and any associated unintended consequences were examined using LTM. This focused on how intervention effects were impacted by three important factors that are not easily controlled by researchers:

- the contexts in which the intervention is being developed, implemented and maintained;

- variations between people and places and across time in how the intervention is implemented;
- the participants' experiences of and perspectives on the intervention.

A process evaluation (PE) checklist was developed based on the work of Nielsen and Randall (2013). This was reviewed by the full project team to assess the suitability of the questions for the Tideway case. This identified questions that required no changes, questions requiring amendment and questions that needed more radical adaptations to meet the needs of the tracer study and / or the intervention and its context.

-Insert Table 1 about here-

The checklist was reviewed in terms of the applicability of the data collection framework to the interventions (in LTM terms, the tracers); the ease with which stakeholder could provide reliable and valid answers to the questions; and, the availability of evidence one or more sources (the tags). Data collection and analysis is then focused on how the interventions were carried out and identifying what worked, for whom and under what conditions. Three interventions (or 'tracers') were selected to pilot the PE table questions, 'Right Start', PPE and Welfare. Right Start was aimed at improving project start up OSH accident/incident statistics by raising awareness at all levels of the risk of accidents and incidents when sites are being mobilised and employees are unfamiliar with site environments and procedures. It was selected as it focussed on site-based interventions and thus contains potential between-site variations in intervention practices. The PPE initiative covered the roll out of innovative PPE project-wide. Intervention practices and perceptions of them may vary across intervention contexts. The Welfare intervention covered the

implementation of ‘transformational’ welfare facilities compared to existing standards in construction and aimed at improving working conditions to impact on employees’ health and to change attitudes towards working in construction. All three interventions provided potential for emergent and unrealised strategies and for significant variations in participants’ perceptions of the interventions: they were delivered to a diverse workforce working in a variety of operational contexts.

At this stage, the testing of the PE table for each of the selected interventions is still in progress. This involves collecting the additional information to complete a process evaluation framework for each intervention. A comparison will then be made between each intervention as planned and as practiced and through this comparison the reasons for success or failure can be documented. In terms of the PE approach the completed table would then be reviewed by the research team to evaluate the credibility of the data and of the causal pathways identified through its analysis. Context-mediator-outcome pathways will then be established: these describe how the intervention setting (context) prompts, supports or hinders the delivery of intervention activities (the mediators) and thus changes in workers’ thinking and behaviour (outcomes). These findings inform modifications to on-going interventions (e.g. by bolstering unrealised strategies or setting the conditions for effective emergent strategies and practices). The extent to which the findings resonate with other intervention settings will then be assessed through consultation with OSH practitioners, the client organisation(s) and other stakeholders in the form of lessons learnt and/or implementation guidance.

Discussion

Megaprojects evolve over long time periods often with unpredictable outcomes. The reasons for these outcomes are complex being shaped by interconnected factors and

intervention contexts during intervention initiation, design, delivery and modification processes throughout the lifespan of the project. It is not uncommon for strategy to practice links to remain unfulfilled or to develop in unexpected ways and for them be buffeted and shaped by external events (Jarzabkowski, Kaplan, Seidl, & Whittington, 2016; Mirabeau, Maguire, & Hardy, 2018).

Realist evaluation using LTM takes account of these factors. Such an approach unpacks and examines in detail the complex causal pathways of intervention outcomes. It provides data about how and why interventions work in situ and often these are unopened black boxes in intervention evaluation carried out using other methods. In doing so, the approach makes evaluation significantly more complex but also more resource efficient, informative and useful (for example by identifying problems with the intervention that can be rectified).

Nielsen and Randall's (2013) model was designed to be flexible enough to capture data on a range of issues not all of which will be relevant to LTM throughout the lifecycle of every intervention project. In applying it to Tideway only relatively modest changes were needed to make the checklist fit for use across a number of different OSH interventions. The model also provides limited information about the history that led up to the project beyond some rudimentary information about the initiation process. LTM could be expanded to include more information about similar interventions participants may have experienced in the past. Realist evaluation should be developed from 'what is already known' (Pawson, 2013, p. xvi) about the issue. This may be especially important when the intervention has led to inconsistent and disappointing outcomes in the past.

LTM requires the collection of large amounts of data with relatively high frequency. However, the method is made efficient through the use of data routinely

collected by stakeholders. The Right Start, Welfare and PPE intervention will use data routinely collected and used for a variety of purposes within the organisations involved, as well as a relatively small amount of additional data collected by the researcher team. The PE checklist can be used in four ways: to identify where there are gaps in the researchers' knowledge; to identify where the missing information can be obtained; to provide a consistent approach to evaluation across interventions and contexts; and, to facilitate the identification of data that will enable the triangulation of findings. Where additional data are required, low-cost and minimally disruptive data collection techniques are available (e.g. focus groups, brief interviews and short questionnaires). The main benefit of LTM is that the focus is on the process of change initiated by the intervention from three viewpoints; the overall context, the observable intervention activities and the views of the people involved/affected. These data can be used to pin-point reasons for intervention success and failures. This is made possible by a longitudinal design that allows the changes and their weaving patterns to be studied from inception through implementation and finally embedded in practice.

Overall, the study was used to reveal new approaches to achieving desirable OSH outcomes, together with in-depth knowledge of how they can best be managed through the process of implementation. The research focused on a number of areas including: leadership, strategies, standards and professional practices; achieving transformational OSH performance; OSH knowledge, expertise and innovation; policy to practice translation/realisation. A key aim of the ongoing research is to identify the practical lessons, knowledge and good practice that are developed and share these with wider industry. Exploring how this is achieved will provide a response to the call

for SaP research that addresses the macro as well as the micro perspectives of strategy implementation (Jarzabkowski & Spee, 2009; Seidl & Whittington, 2014).

However, LTM is not without its challenges. The identification of suitable 'tracers' and 'tags' within each intervention and context is not straightforward and requires the researchers to work with stakeholders to gather up local intelligence about the intervention strategy, activities and contexts. In the evaluation of a mega-project it involves managing data organised and handled by many different stakeholders. It also requires very good working relationships between the researchers and the megaproject employees. This was facilitated on Tideway by having four researchers embedded into the project teams of Tideway and the three main joint venture main works contractors over the three years, such that they were treated very much as normal employees.

Ethical issues associated with data collection can be complicated by the unpredictability of the need for, and relevance of, different types of data.

Arrangements for the collection of important data may need to be enacted with little advance warning as intervention activities change and unfold. Organising and coding data can also be challenging as researchers may not be able to easily anticipate strategy in practice activities (and how these differ from strategies as planned). The intervention examples we have described are from an early stage LTM study and as such the data collection process is likely to change over time: realist approaches are flexible enough to be adapted to developments in intervention theory and evaluation methodology. This allows for different approaches to be sought and tested. Data analysis may become more time-consuming as a result but will be better aligned with intervention delivery and not restricted to the protocols linked to the intervention plan.

Conclusion

A methodology that is flexible in terms of opening up and shutting down lines of inquiry can make better uses of existing resources to evaluate large and complex organisational change. LTM, which includes process evaluation, provides a resource efficient and rigorous approach to the evaluation of complex interventions. It can provide a better understanding of decision-making that affects the continuous improvement and transfer of OSH intervention activities in complex construction environments.

References

- Baccarini, D. (1996). The concept of project complexity: A review. *International Journal of Project Management*, 14, 201-204. doi: 10.1016/0263-7863(95)00093-3
- Bryman, A. (2012). *Social research methods (4th ed.)*. Oxford: Oxford University Press.
- Chau, V. S., & Witcher, B. J. (2005). Longitudinal tracer studies: Research methodology of the middle range. *British Journal of Management*, 16, 343-355. doi: 10.1111/j.1467-8551.2005.00459.x
- Combey, P. (1980). A tracer approach to the study of organizations. *Journal of Management Studies*, 17, 96-126.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Davies, A., & Mackenzie, I. (2014). Project complexity and systems integration: Constructing the London 2012 Olympics and Paralympics Games. *International Journal of Project Management*, 32, 773–790. doi: 10.1016/j.ijproman.2013.10.004
- Flyvberg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk: An anatomy of ambition*. Cambridge: Cambridge University Press.
- Fridrich, A., Jenny, G. J., & Bauer, G. F. (2016). Outcome expectancy as a process indicator in comprehensive worksite stress management interventions. *International Journal of Stress Management*, 23, 1-22. doi:10.1037/a0039202

- Fuller, P., Gibb, A., Jones, W., Dainty, A., Haslam, R., Bust, P., & Pinder, J. (2017). Is the long-bow better than the cross-bow? Emerging issues from mobilising a longitudinal study on a mega-project. *Journal of Construction Project Management and Innovation*, 7, 2054–2065. <http://hdl.handle.net/10520/EJC-c2417cc42>
- Griffiths, A. J. (1999). Organizational interventions: Facing the limits of the natural science paradigm. *Scandinavian Journal of Work, Environment and Health*, 25, 589–96. doi: 10.5271/sjweh.485
- Hale, A. (2014). Construction Safety management: Do we know what works? Paper presented at *Achieving Sustainable Construction Health and Safety, CIB W099 International Conference*, Lund, Sweden. Abstract retrieved from <https://lup.lub.lu.se/search/ws/files/6285840/8030501.pdf>
- Harrison, J., & Dawson, L. (2016). Occupational health: Meeting the challenges of the next 20 years. *Safety and Health at Work*, 7, 143–149. doi: 10.1016/j.shaw.2015.12.004
- Hasson, H., von Thiele Schwarz, U., Nielsen, K., & Tafvelin, S. (2016). Are we all in the same boat? The role of perceptual distance in organizational health interventions. *Stress and Health*, 32, 294-303. doi: 10.1002/smi.2703
- Havermans B. M., Schelvis, R. M. C., Boot, C. R. L., Brouwers, E. P. M., Anema, J. R., & Van der Beek A. J. (2016). Process variables in organizational stress management intervention (SMI) evaluation research: A systematic review. *Scandinavian Journal of Work Environment and Health*, 42, 371-381. <http://dx.doi.org/10.5271/sjweh.3570>

- Hornby, P., & Symon, G. (1994). Tracer studies. In C. Cassell & G. Symon (Eds.) *Qualitative methods in organisational research: A practical guide* (pp. 167-186). Thousand Oaks: Sage.
- Jarzabkowski, P. (2004). Strategy as practice: Recursiveness, adaptation, and practices-in-use. *Organization Studies*, 25, 529-560.
<https://doi.org/10.1177/0170840604040675>
- Jarzabkowski, P., Kaplan, S., Seidl, D., & Whittington, R. (2016). On the risk of studying practices in isolation: Linking what, who, and how in strategy research. *Strategic Organisation*, 14, 248-259.
[doi:10.1177/1476127015604125](https://doi.org/10.1177/1476127015604125)
- Jarzabkowski, P., & Spee, A. P. (2009). Strategy as practice: A review and future directions for the field. *International Journal of Management Reviews*, 11, 69-95. doi: 10.1111/j.1468-2370.2008.00250.x
- Johns, G. (2006). The essential impact of context on organizational behavior. *The Academy of Management Review*, 31(2), 386-408. doi: 10.2307/20159208.
- Johns, G. (2018). Advances in the treatment of context in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 21-46. doi: 10.1146/annurev-orgpsych-032117-104406.
- Miller, D., & Friesen, P. H. (1982). The longitudinal analysis of organizations: A methodological perspective. *Management Science*, 28, 1013-1034. doi: 10.1287/mnsc.28.9.1013
- Mintzberg, H. (1987). The strategy concept 1:5 Ps for strategy. California *Management Review*, 30, 11-24. doi: 10.2307/41165263

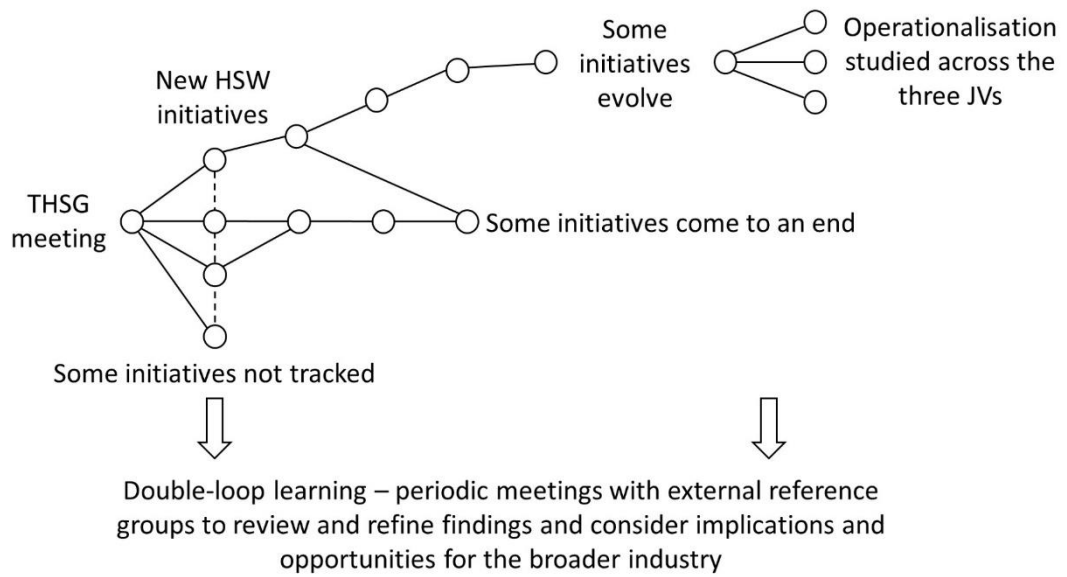
- Mirabeau, L., & Maguire, S. (2014). From autonomous strategic behavior to emergent strategy. *Strategic Management Journal*, *35*, 1202–1229. doi: 10.1002/smj.2149
- Mirabeau, L., Maguire, S., & Hardy, C. (2018). Bridging practice and process research to study transient manifestations of strategy. *Strategic Management*, *39*, 582-605. doi: 10.1002/smj.2732
- Nielsen, K. (2017). Organizational occupational health interventions: What works for whom in which circumstances? *Occupational Medicine*, *67*, 410–412. doi:10.1093/occmed/kqx058
- Nielsen, K., & Abildgaard, J. S. (2013). Organizational interventions: A research-based framework for the evaluation of both process and effects. *Work and Stress*, *27*, 278-297. doi: 10.1080/02678373.2013.812358
- Nielsen, K., & Miraglia, M. (2017). What works for whom in which circumstances? On the need to move beyond the ‘what works?’ question in organizational intervention research. *Human Relations*, *70*, 40–62. doi: 10.1177/0018726716670226
- Nielsen, K., & Randall, R. (2013). Opening the black box: A framework for evaluating organizational-level occupational health interventions. *European Journal of Work and Organizational Psychology*, *22*, 601-617. doi: 10.1080/1359432X.2012.690556
- Nielsen, K., Randall, R., Holten, A. L., & González, E. R. (2010). Conducting organizational-level occupational health interventions: What works? *Work and Stress*, *24*, 234–259. doi: 10.1080/02678373.2010.515393

- Nytrø, K., Saksvik, P. Ø., Mikkelsen, A., Bohle, P., & Quinlan, M. (2000). An appraisal of key factors in the implementation of occupational stress interventions. *Work and Stress, 14*, 213-225. doi: 10.1080/02678370010024749
- Pawson, R. (2013). *The science of evaluation: A realist manifesto*. Sage: London.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: SAGE.
- Pawson, R., & Manzano-Santaella, A. (2012). A realist diagnostic workshop. *Evaluation, 18*, 176–191. doi: 10.1177/1356389012440912
- Pettigrew, A. M. (1992). The character and significance of strategy process research. *Strategic Management Journal, 13*, 5-16. <https://www.jstor.org/stable/2486363>
- Pettigrew, A. M. (1990). Longitudinal field research on change: Theory and practice. *Organization Science, 1*, 267-292. doi:10.1287/orsc.1.3.267
- Randall, R. (2013). Process monitoring in intervention research: A 'dashboard' with six dimensions. In G. F. Bauer & G. J. Jenny (Eds.), *Salutogenic organizations and change* (pp. 259-274). London: Springer Science and Business.
- Randall, R., Griffiths, A., & Cox, T. (2005). Evaluating organizational stress-management interventions using adapted study designs. *European Journal of Work and Organizational Psychology, 14*, 23-41. doi: 10.1080/13594320444000209
- Randall, R., Nielsen, N., & Tvedt, S. (2009). The development of five scales to measure participants' appraisals of organizational-level stress management interventions. *Work and Stress, 23*, 1-23. doi: 10.1080/02678370902815277
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*, 29-48. doi: 10.1177/1356389007084674

- Rossi, P. H., Lipsey, M. W., & Freeman, H. E (2003). *Evaluation: A systematic approach (7th ed.)*. Thousand Oaks, CA: Sage.
- Sherratt, F. (2018). Shaping the discourse of worker health in the UK construction industry. *Construction Management and Economics*, 36, 141-152. doi: 10.1080/01446193.2017.1337916
- Seidl, D., & Whittington, R. (2014). Enlarging the Strategy-as-Practice research agenda: Towards taller and flatter ontologies. *Organization Studies*, 35, 1407–21. doi: 10.1177/0170840614541886
- Tsoukas, H., & Chia, R. (2002). On organizational becoming: Rethinking organizational change. *Organization Science*, 13, 567-582. doi: 10.1287/orsc.13.5.567.7810
- van der Molen, H. F., Basnet, P., Hoonakker, P. L. T., Lehtola, M. M., Lappalainen, J., Frings-Dresen, M. H. W., Haslam, R., & Verbeek, J. H. (2018). *Interventions to prevent injuries in construction workers*. Cochrane Database of Systematic Reviews Art. No.: CD006251. Retrieved from <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD006251.pub4/epdf/abstract>. DOI: 10.1002/14651858.CD006251.pub4.
- von Thiele Schwarz, U., Lundmar, R., & Hasson, H. (2016). The Dynamic Integrated Evaluation Model (DIEM): Achieving sustainability in organizational intervention through a participatory evaluation approach. *Stress and Health*, 32, 285-293. doi: 10.1002/smi.2701
- Waterman L. (2007) The future for occupational health in the United Kingdom. *Policy and Practice in Health and Safety*, 5:sup1, 125-140. doi:10.1080/14774003.2007.11667707

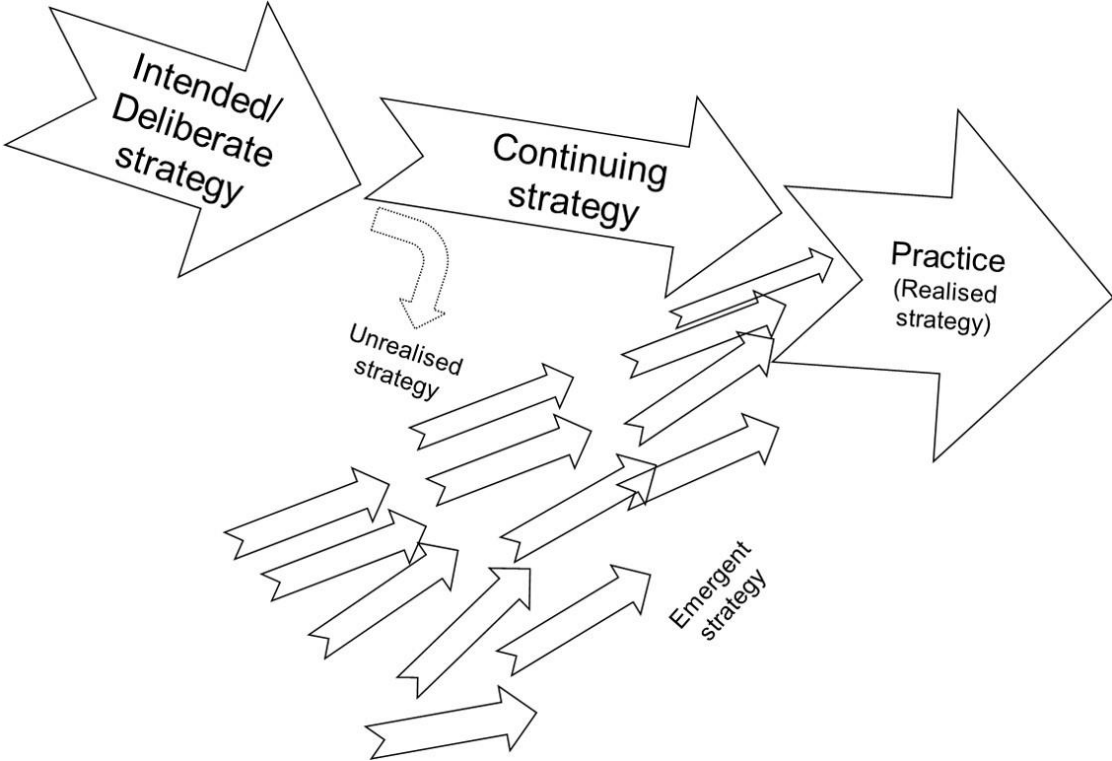
- Weiner, B. J. (2009). A theory of organizational readiness for change. *Implementation Science*, 4, 67. doi: 10.1186/1748-5908-4-67
- Whittington, R. (2006). Completing the practice turn in strategy research. *Organization Studies*, 27, 613-634. doi: 10.1177/0170840606064101
- Whysall, Z., Haslam, C., & Haslam, R. (2006). A stage of change approach to reducing occupational ill health. *Preventive Medicine*, 43, 422-428. doi: 10.1016/j.ypmed.2006.07.004
- Woodward, J. (1970). *Industrial organization: Behaviour and control*. Oxford: Oxford University Press.
- Yin, R. K. (2003). *Case study research: Design and methods (3rd ed.)*. Thousand Oaks, California: Sage.

Figure 1. Evolution of HSW initiatives in the case study project



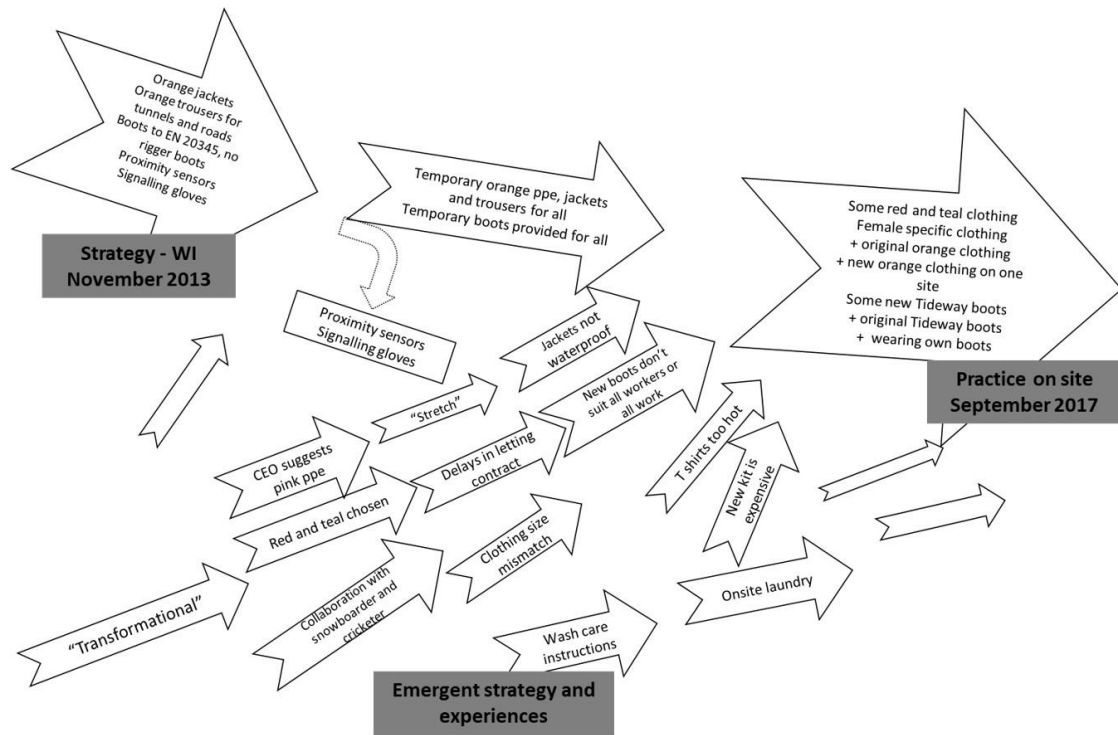
Note. THSG = Transformational Health Safety Group; HSW = Health Safety and Welfare;
MWC = Main Works Contractors.

Figure 2. Adapted version of Mintzberg's model of strategy.



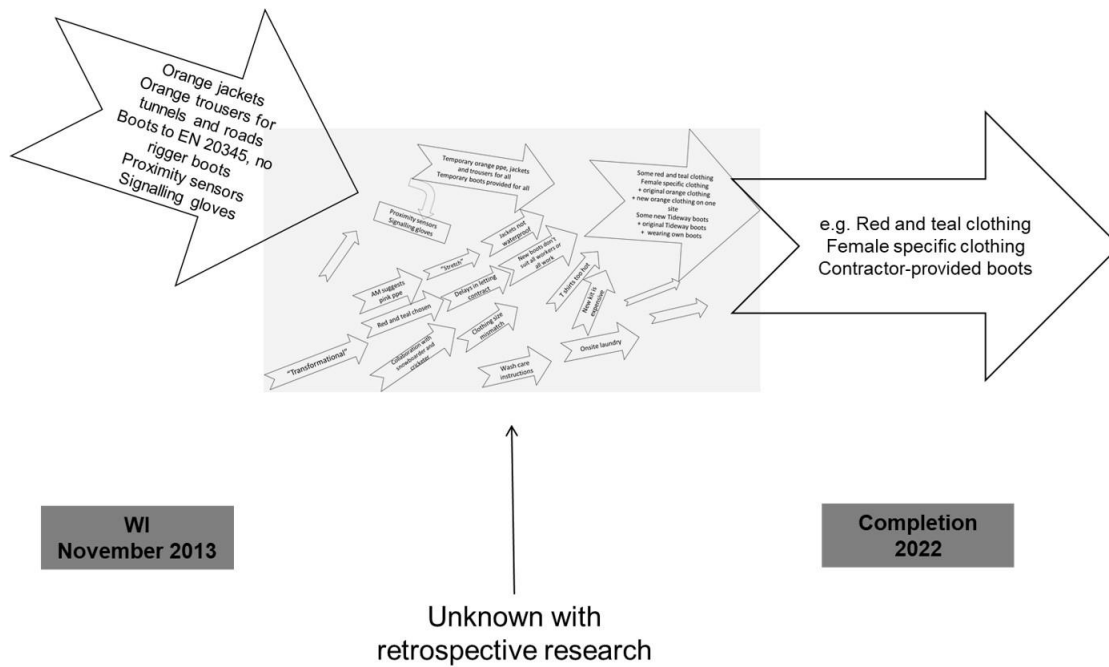
Note. Adapted from Mintzberg, H. (1987). The strategy concept 1:5 Ps for strategy. *California Management Review*, 30, 11-24.

Figure 3. Understanding how strategy informs practice in real time, through longitudinal research.



Note. WI = Works Instructions

Figure 4. Limitations of cross sectional or retrospective research.



Note. WI = Works Instructions

Table 1. Modifications made to the Nielsen and Randall (2013) process evaluation model when applied to Tideway OSH interventions.

Original	Modified version
Participants mental models	Participants knowledge, experiences, perceptions of, and attitudes to, intervention activities?
What is the role of participants' mental models?	What is the role of participants' mental models (see above) in shaping intervention experiences and outcomes?
To which extent are participants ready for change?	To what extent do participants indicate that they are ready for the intervention activities?
To which degree do participants have shared mental models? In case of divergence, how did mental models differ?	To what degree do participants see the intervention aims and activities in a homogenous / heterogeneous way? In case of divergence, how did views of the intervention differ? What are the main variations between participants and within participants over time? What are the implications of divergence?
Did the intervention bring about a change in participants' mental models?	Did the intervention bring about a change in participants' knowledge, skills and attitudes?
Why were intervention activities not implemented?	According to key stakeholders why were intervention activities: implemented / not implemented? Sustained / not sustained? Modified / not modified? What new strategies emerged and were these translated into practice?