DATA DRIVEN PREDICTIVE RISK MODELLING

Marcus Beard, BEng

Arthur D. Little

SUMMARY

This paper presents the work that Arthur D. Little (ADL) completed to develop a new safety risk model for a major utilities sector client. The risks in question are low probability, high consequence risks, arising from safety critical lone workers leaving unsafe conditions in their work. This new model depends on using multiple disparate data sources effectively to provide indicators of employees potentially at risk of leaving these unsafe conditions. The model takes diverse performance measures to contribute to a reliable risk metric that informs management decision making in both the operational and safety assurance functions.

ADL used a data driven approach to predict each worker's risk, using diverse sources of information to build and validate the model. The type of modelling completed is applicable to many industries, including the rail sector, where there are significant, and growing, quantities of data available and a diverse asset base. The approach to model development focussed on turning large volumes of data across a population of several thousand employees into actionable management information. This information then allows managers to take decisions on rectifying these unsafe conditions through focussing resources.

This modelling approach is well suited to the rail industry, which often has large geographically dispersed asset bases (many of these assets having long lives), significant lone working, a mixture of recent and experienced employees, as well a large amount of quantitative data available. The principles of the model we developed are directly transferable to the rail industry and support more effective and efficient safety risk management.

INTRODUCTION

ADL developed a new, data driven risk model for a utilities sector client who needed a way to better prioritise the time their safety assurance resources spent inspecting large volumes of safety critical work completed by lone workers over a large, geographically spread area. These workers can sometimes leave unsafe situations in their work that can lead to low probability, high severity accidents with significant reputational risk. We modelled the risk of each individual worker leaving unsafe situations (referred to as "defects" in the rest of this paper) using a composite function of multiple independent parameters, such as driving behaviour, experience and productivity, found through statistical analysis.

We used the analysis to develop a risk model that outperformed the client's existing risk model and supported their efforts to refocus both safety assurance and operational management attention to where it was most useful. The company has adopted the model to help their operational and assurance managers further control the risk of these low probability, high consequence events.

DATA USE IN INFORMING MANAGEMENT ACTION

Companies in many sectors, including the rail sector, now have access to large volumes of previously unavailable data that, used in the right way, could provide indications of various aspects of risk. Typically, data is distributed across multiple databases, with no specific individual having visibility over everything. This data could relate to asset condition, previous incidents, employee performance, or any other aspects of a business. But data is just data. It is turning this data into information that allows managers to inform more robust decisions that is increasingly important to improve productivity and risk management. Sitting on large quantities of data without taking action can be problematic (from a reputational as well as a legal perspective) in the event of an accident, as it might be judged that there was knowledge that was not acted upon. Such 'foreseeability' is often used as an argument in prosecutions.

Variable data quality is also a challenge in many businesses – our experience shows that it is common for two different databases containing employee data to show different values or for one database to be out of data or incomplete. For example, one database might show employee productivity and another might show driving telematics data for these employees, but some employees may be omitted from one or both databases. In our experience even large companies struggle to maintain consistency across these disparate databases, and this can be a symptom of managers in different functions working in silos, with limited cross-functional engagement.

Data that relates to equipment failures and human errors often relate to events with low probabilities, but potentially catastrophic consequences (as was the case with our particular assignment that is the background to this paper). Physical assets that are spread across a wide geographic area, make comprehensive assessments or checks impracticable. Unsupervised working is common: an employee installing a new asset who does not correctly follow a procedure may leave a latently unsafe situation that will go unchecked for most or all of its life.

Companies are starting to use broader, indirect data to predict which unsupervised workers are more likely to make errors or miss out steps in their work. Converting this data into information such that manages can focus their attention, such as pre-emptive maintenance or additional training, is a challenge. Translating data used in the risk model into information that managers can use to take decisions is critical – managers need guidance on what the model outputs mean to them. The results of the modelling should be clear and simple to understand, and appropriate for managers at all organisational levels. Both operational and safety/assurance functions should take actions based on the results of the risk modelling to ensure risks continue to be controlled (Figure 1).

Figure 1: Operations and assurance functions must work together to turn data into actionable management information

Our method of implementing a data-driven model such as this is follows a four-step approach:

- Review available data from databases across multiple business functions
- Analyse correlations between data and the undesired acts or events to find those data parameters with strong predictive ability
- Develop and validate a risk model based on these correlations
- Use the model outputs to drive management actions, such as:
  - Focussed monitoring by safety or assurance functions for individuals at risk of leaving unsafe conditions, but who have not yet left them
  - Additional support from operational management for known higher-risk individuals
  - Increased monitoring and inspection of physical assets at risk of developing unsafe conditions

This model development process can also highlight the challenges raised earlier in this section. For example, it can often reveal inconsistencies between databases, highlighting the value to be gained by operational and support functions working together, rather than in separate silos.

PROJECT BACKGROUND

Our client was a large European utility company's metering business, responsible for fitting both smart and "dumb" electricity and gas meters. They replace and install hundreds of thousands gas and electricity meters every year, relying on a sample-based inspection approach to gain assurance that work is being performed correctly.

During 2016 they completed an audit of their safety assurance process and found that their actual defect rate was 11 times higher than that indicated by their own safety assurance process. We reviewed their audit methodology and found that it was robust, showing that they needed a new way to prioritise their safety assurance inspection activities to be able to better find meter fitters likely to leave (or have already left) defects in their work. We also reviewed their existing safety assurance processes, finding some underlying management issues that were negatively influencing their technical safety performance.

The client incorporated our findings and recommendations into a larger ongoing project and requested ADL's support in developing a new risk model to allow them to better prioritise their use of safety assurance resources. The model needed to provide a better predictive capability than their existing safety assurance regime (primarily based on length of service), to identify both meter fitters that were likely to have already left defects, and those likely to leave defects in the future.

INDUSTRY CONTEXT

Many governments have established ambitious targets for smart electricity meter fitment, for example, the European Union (EU) has set a target of 80% by 2020. These targets are driven by smart grid strategies that aim to improve the reliability and efficiency of energy supplies, whilst reducing costs. These stretching targets need a clear, significant response from suppliers, to meet the deadlines set.

Suppliers and distributors have responded by rapidly recruiting and training large numbers of meter installers to provide the necessary installation capacity, which has disrupted businesses that had been (in many cases) operating in a steady state for long periods.

Replacing and installing utility meters can introduce additional risks to people and property – for example, electrical wiring defects introduced during installation can cause overheating, fire or electrocution, and gas leaks can also lead to serious fires or explosions. The three main electrical wiring faults are:

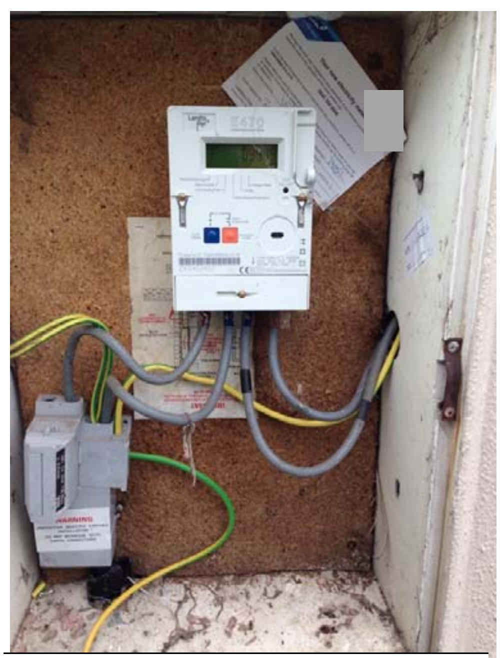

Figure 2: Example of an electricity meter (top) and cut-out fuse (bottom-left)

- Cross/reverse polarity (live and neutral cables connected to the wrong terminals)
- Loose connections at the meter terminals
- Exposed copper at either the meter or cut-out fuse (see Figure 2)

These types of defects (known as high risk technical defects) may not always be visible from a visual inspection and can be present for long periods without any disruption to a customer's supply being noticed, effectively acting as a latent risk. This leads to the potential for a significant impact on a company's reputation in the event of an incident due to the residential nature of many locations.

Based on our experience, at least 1% of all electricity and gas meters have been installed with one or more high-risk technical defects. Historic incident reports show that the probability of any individual hazardous situation developing realising the potential consequences of the hazard is small. However, when this is considered across the installed utility meter base there are potentially tens of thousands of incorrectly installed meters in operation with latent hazards.

CAUSES OF DEFECTS

Installing or changing a meter is a well-defined, straightforward task. The meter fitters receive extensive training in the role and are accompanied by a mentor once this training is over, with a requirement for their competence to be signed off. However, once this mentoring process is complete and they are working on their own, meter fitters begin to leave defects (see Figure 3). This analysis also showed that, among those fitters who had not left defects after their first two years, the chance of any of them leaving a defect decreases over time, demonstrating the value of experience in working safely in this context.

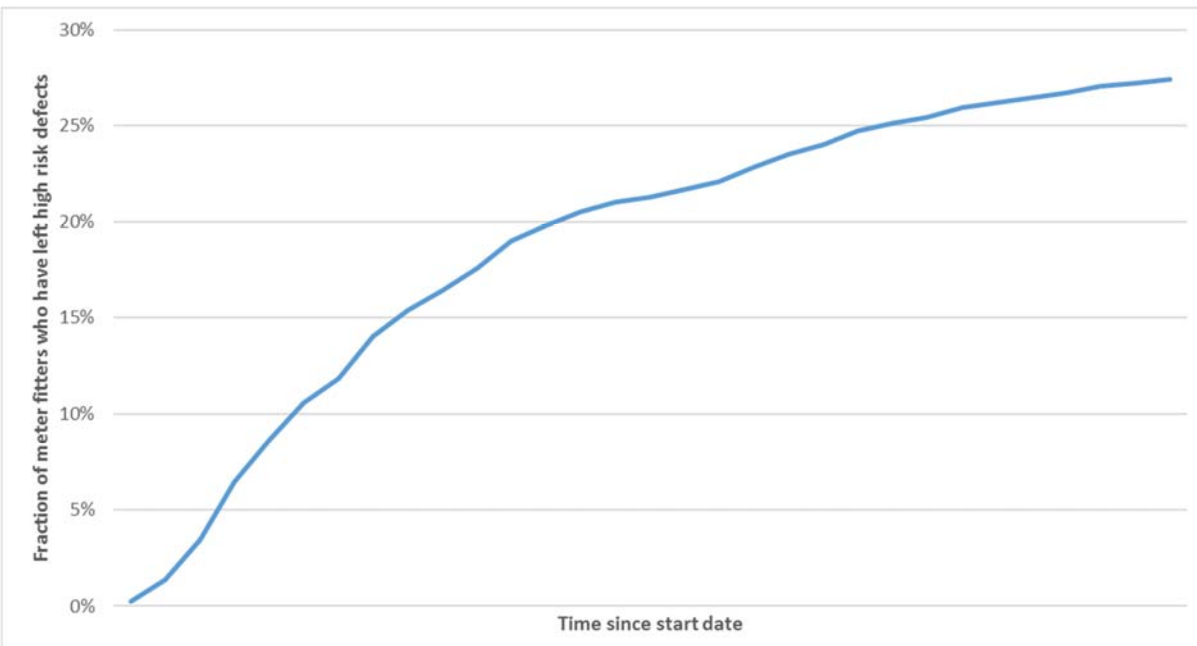

Figure 3: Cumulative fraction of meter fitters who have left defects since their start date

These defects are generally caused by behavioural issues and are not due to gaps in meter fitters' competence. Fitters either skip steps in the procedure to complete the task more quickly, or forget to perform certain steps, such as checking that all screws are fully tightened once the new meter has been fitted.

This behavioural driver for leaving defects in work completed provided a starting point for our analysis and development of the new risk model. Showing correlations between meter fitters who have left defects in their work and other behavioural indicators, such as driving telematics scores, would provide a statistically-based approach to predicting which meter fitters who had not yet left defects might be likely to in the future.

NEW RISK MODEL DEVELOPMENT

The initial basis for the new risk model was a two by two matrix, assigning each meter fitter to one of four categories (Figure 4), with one axis showing a *Recorded Risk* and the other axis a *Modelled Risk*. We then further developed this into a 10 by 11 matrix in the final model to provide a higher resolution risk classification.

Figure 4: Categorisation of meter fitters in the new risk model

Analysis from our review of the client's safety assurance audit showed that approximately 45% of all high-risk defects were left by meter fitters who had not left a high-risk defect – the new model therefore could not only rely on high-risk technical defects as a predictor of future, but had to include other parameters.

The *Recorded Risk* axis used the client's existing safety risk measure and is based on observed defects that their safety assurance function has found, rescaling the scores to a value between 0 and 10.

The *Modelled Risk* axis we proposed is a composite measure based on our analysis of factors that had a demonstrable correlation with defect rates. Developing the Modelled Risk axis was a four-step process:

- From our analysis we found a range of parameters that correlate with meter fitters who are more likely to leave high risk defects in gas and electrical meter installations
- For each parameter high, moderate and low risk boundaries for relative risk were defined
- We then weighted these parameters by assigning values to the different values each parameter can take, based on the strength of the correlation and its influence on defect rates.
- Testing and validation of the model

The analysis of parameters that correlate with meter fitters who are more likely to leave high-risk defects in gas and electrical meter installations showed seven parameters that had a strong or moderate level of correlation with leaving high risk defects (Figure 5).

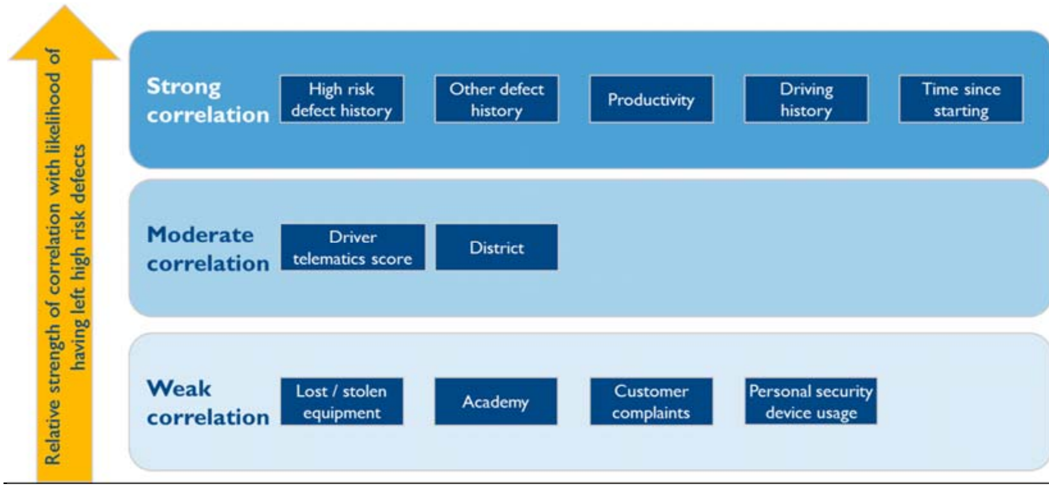

Figure 5: Strength of correlation of parameters with meter fitters who have left high-risk defects

Each individual parameter was broken down into three categories of relative risk to which meter fitters were allocated against boundary values (Figure 6). We completed this weighting by splitting the range of quantitative values into low, moderate and high-risk categories based on quantiles of meter fitters, so that for each parameter less than 20% of meter fitters would be allocated to the high-risk category.

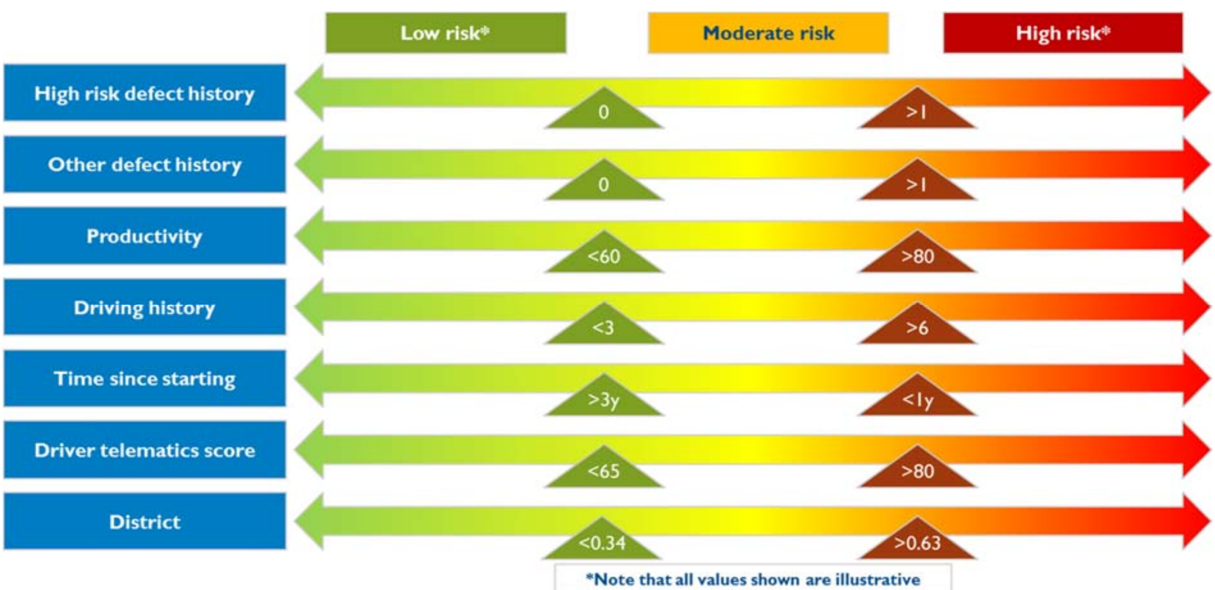

Figure 6: Definition of low, moderate and high-risk boundaries for the seven Modelled Risk parameters

We then weighted these seven parameters by assigning values to moderate and high-risk fitters for each parameter (Figure 7), with a fitter's overall Modelled Risk being the sum of their individual parameter weightings. The combined sum of the individual parameter weightings (i.e. their Modelled Risk) then provides an indication of the chance that a fitter may leave a defect in the future.

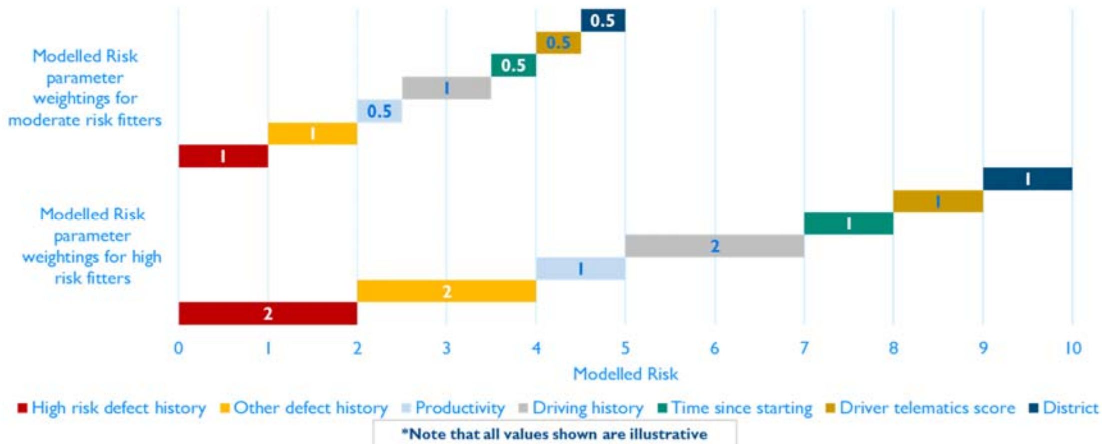

Figure 7: Parameter weightings assigned to fitters with moderate and high risks for each individual parameter

These weightings were then adjusted through sensitivity analysis to achieve a distribution of Modelled Risk among fitters where 20% were allocated to the top half of the matrix in Figure 4. We completed this sensitivity analysis to ensure that the new risk model differentiated sufficiently between the fitters with the greatest risk factors, so that the client could focus their safety assurance and operational management attention where it would provide the greatest benefit.

We then validated the model's effectiveness by comparing the distribution of fitters on the new risk model matrix (Figure 4) who had left defects after we had completed the model development, with those fitters who had not left defects in the analysis period under two different scenarios (Figure 8), using an independent data set.

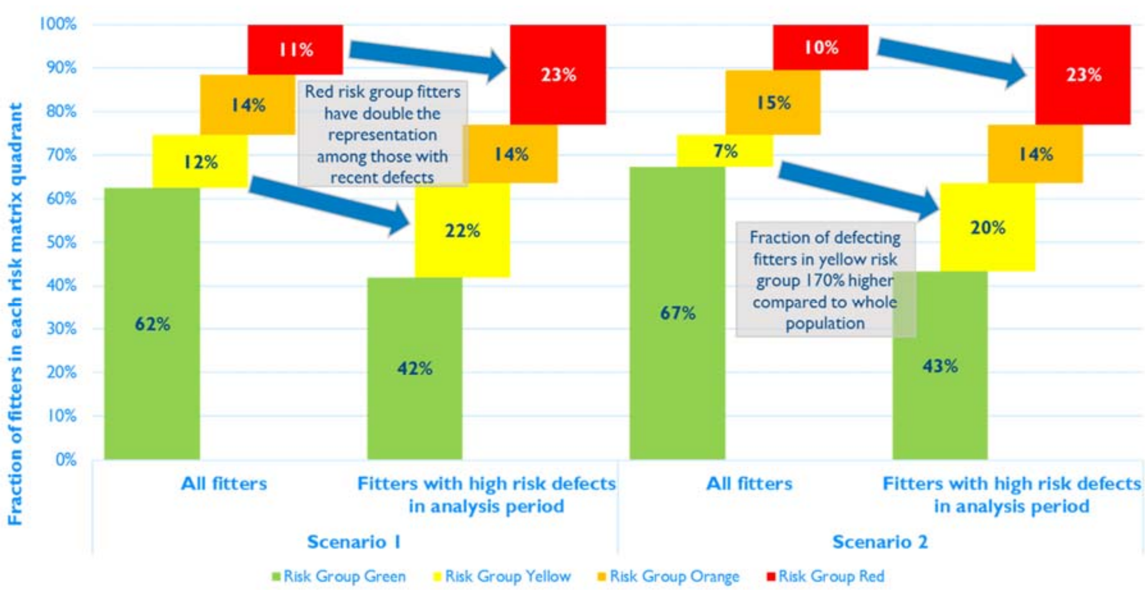

Figure 8: Comparison of distribution of fitters with recent high risk defects against those without defects

We also compared the performance of the new risk model with the client's previous risk model and found that it was approximately twice as effective in finding fitters likely to leave defects in the future, without producing too many false positive results.

The client is currently operationalising the risk model, having completed consultation with management and trade unions on its use.

POTENTIAL APPLICATIONS IN THE RAIL INDUSTRY

The new risk model gave our client an enhanced basis to model which fitters are more likely to leave defects in their work, allowing them to focus limited safety assurance resources on those individuals at greater risk of leaving hazardous situations. Whilst the individual risk posed by an incorrectly installed meter is very low, the high number of installations, geographic spread, and unsupervised nature of the job mean that, collectively, the risk is not negligible.

This type of modelling is relevant to the rail industry as many assets are similar to those in utility companies, for example:

- Assets can be spread over a wide geographic area in uncontrolled environments (e.g. bridges, permanent way, earthworks, trackside signalling equipment)
- Assets often have useful lives measured in decades and can go long periods between inspections
- For some assets, inspections may not consistently detect certain types of latent defects or problems
- The delivery capacity of work can be much higher than the resources available to inspect that work has been completed properly

The railway industry also has a significant amount of data available, relating to both assets and people, which can be spread across different departments and is not being used to full effect at present. This data contains information that could be used to further develop risk-based approaches to either asset inspection or employee/contractor safety performance. This would be consistent with the move away from time-based or sample-based inspection approaches that has occurred in some railways over recent years.

A similar approach to risk modelling has already been used in the GB railway for many years to model the risks at level crossings. This is a quantitative model based on large volumes of data, collected from level crossing censuses over many years, which models both the likely frequency and potential consequence of any level crossing incident. The calculations use both historic incident records and specific level crossing parameters, such as sighting distance and usage levels, to predict the likely level of risk at any level crossing in the country. The model has been successful in allowing Network Rail to prioritise investment in improving level crossing safety, with significant reductions in the overall national level of level crossing in recent years.

The most applicable areas that this quantitative, data driven approach can be used in the railway industry are likely to be those where:

- The nature of the work being completed makes it hard to check (e.g. due to geographic spread or number of assets), and defects are not easily found from a visual inspection
- Behavioural factors play a large role in the development of unsafe conditions, where workers missing out steps of a procedure
- Work completed is not always checked by third parties, or where there is a large element of lone working
- There are good quality records of assets, employees and contractors available, even if these are spread across multiple different departments or business functions

Critical to the success of any risk model is the management arrangements that support its use. Clear responsibility and approaches to managing individuals in each category of risk need to reinforce messages that are in line with a company's overall approach to safety management.

CONCLUSION

ADL developed a new risk model for a utility sector client that uses observable and measurable data from a range of sources and databases to predict which specific individual employees are more likely to leave unsafe situations in their work. It combines information from their previous risk model with a new composite measure of risk to give both operational and safety assurance functions the information needed to be able to manage these employees effectively and control the risks associated with a large, geographically dispersed asset base, installed and maintained by lone workers. The model has direct applicability to the rail industry, with infrastructure assets in particular sharing many of these properties.

This data driven approach to risk modelling is possible due to the proliferation of data now available in operational industries and the databases that support it. These types of models rely on high quality input data from databases that are overseen by different departments or business functions and therefore require a collaborative approach between individual managers and moving away from operating in discrete silos. Using data driven risk models to provide managers with actionable management information can lead to improvements in safety performance. These improvements can be made without significant additional expenditure on assurance or inspection resources through making better use of those resources already available to managers in the rail industry.