Modernising operational risk management in financial institutions via data-driven causal factors analysis: A pre-registered report

Cornwell, Nikki; Bilson, Christoper; Gepp, Adrian; Stern, Steven; Vanstone, Bruce

Pacific-Basin Finance Journal

DOI:
10.1016/j.pacfin.2022.101906

Published: 01/02/2023

Modernising operational risk management in financial institutions via data-driven causal factors analysis: A pre-registered report

To enable more proactive management of the underlying sources of operational risks in financial institutions, this pre-registered study seeks to improve traditional qualitative approaches to causal factors analysis. A Bayesian network-based approach is used to leverage both incident and operations data to model the probability of operational loss events. The approach is applied and empirically tested in a case study on an Australian insurance company. The outputs from the model go beyond simply identifying key risk drivers to offer risk managers a deeper understanding of how causal factors influence risk. Insights into the collective effects of causal factors, their relative importance and critical thresholds strategically inform more efficient and effective mitigation decisions, ultimately enhancing firm performance and value.

Keywords: risk management; operational risk; data analytics; firm value; financial institutions; insurance

JEL Classification Codes: G20, G32, D81, C44

1 Introduction

1.1 Brief Background and Context

Operational, or non-financial, risks have led to an average of A$36.7 billion of losses in financial institutions (FIs) globally[1] per annum since 2016 (Operational Riskdata eXchange Association, 2022). While a substantial direct financial impact to FIs, hidden behind this statistic are people (customers and employees) personally impacted, as well as associated reputational damage to brands. Formally defined as "risk[s] of loss[es]

---

[1] Data from the ORX Annual Banking and Insurance Loss Reports based on operational loss events reported by over 100 member firms across the globe. Converted to Australian dollars using the average EUR-AUD exchange rate for the period of 1.57.

resulting from inadequate or failed internal processes, people and systems or from external events" (Basel Committee on Banking Supervision, 2006, p. 144), at their core, operational risks arise from issues in day-to-day business operations – the behaviours and characteristics of customers, the processes performed to service customers, the actions of employees, the performance of technology and systems, and vendors and suppliers.

The digital operating environments of most organisations, and certainly FIs, mean information about many business operations is automatically captured in electronic systems. In other words, the precursors to many operational risks and incidents (i.e., issues in day-to-day operations) are captured in standard operating systems. So, if leveraged with an operational risk mindset, these data streams could reveal the key drivers behind incidents and losses, which mitigation activities can then be targeted toward. While beneficial to understand which levers *can* be pulled, for an organisation with limited resources (money, people and time), more information is needed about which levers *should* be pulled to meaningful inform a proactive ORM strategy.

If the effect that each key driver (or collection of drivers) has on the probability of a loss could be quantified, this would provide objective measures to inform strategic decisions. For example, questions like "what factors in the operating environment have the strongest influence on an operational risk?" could be answered. The answers could practically assist organisations in knowing which controls may be more effective in reducing the likelihood of a loss or where to prioritise monitoring and investment in other mitigation efforts. Beyond this, knowing the specific settings or thresholds of factors that exacerbate an operational risk could enable even more targeted preventative mitigation. Redistributing resources to areas of highest impact would offer valuable efficiency gains and free up risk managers' capacity for more complex and emerging ORM considerations, overall reducing volatility and enhancing firm performance.

1.2 Research Question

This study explores the feasibility and value of such a data-driven tool for causal factors analysis (CFA) of operational risks in FIs, answering the following research question:

How can a financial institution effectively leverage internal operational and incident data to model how causal factors influence the probability of operational risk events, as a tool to improve the effectiveness of risk management, reduce risk exposure and thus enhance financial performance?

The research focusses on functional implementation in the real-world and hence includes an industry-partnered application. The approach is expected to advance traditional qualitative, manual and thus periodic and subjective CFA toward quantitative, real-time, and thus dynamic and objective analysis, offering strategically more valuable outputs.

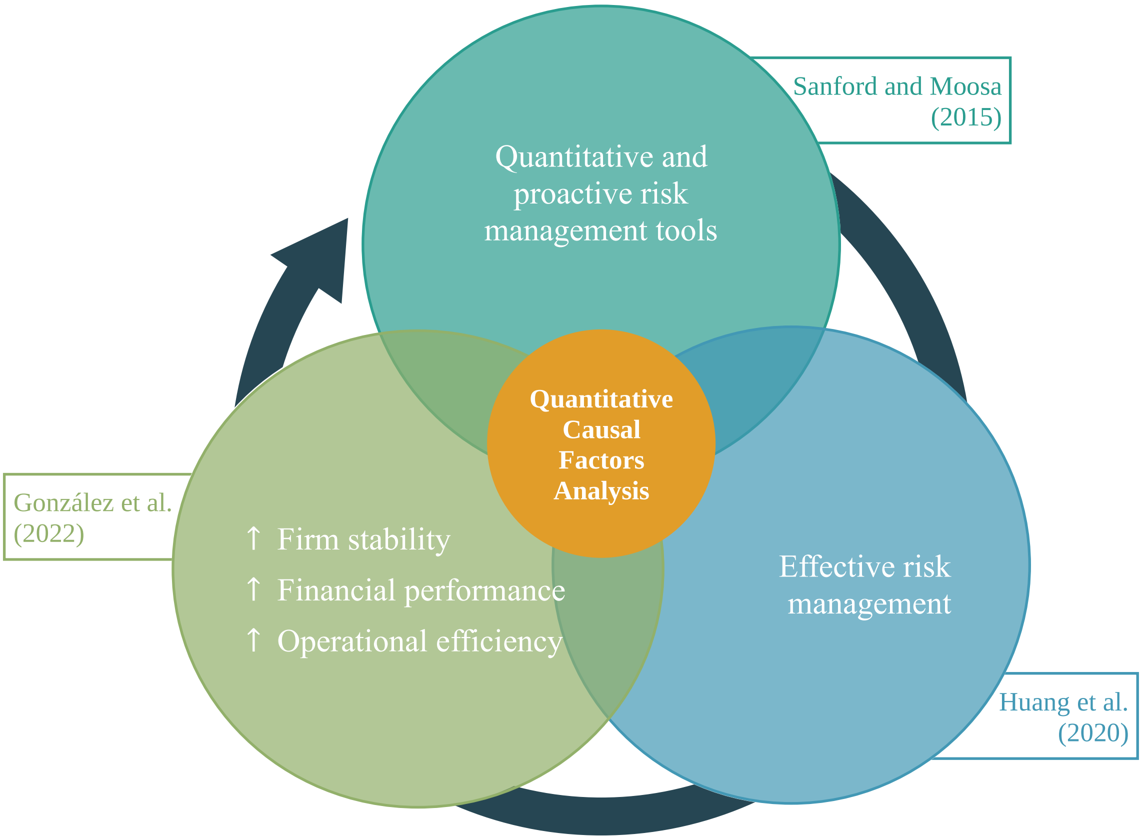
Expanding on the pre-approved research pitches in Appendix A, the remainder of this report details the protocol planned for our "Engagement & Impact" study, constituting Phase 3 of the PBFJ pre-registration publication process. The next section reviews the related literature and motivation of the study (Section 2), culminating in the core idea and hypothesis development (Section 3). The empirical design underpinning the study is outlined in Section 4, including the data and tools for a case study in insurance. Section 5 concludes with the novelties, impacts and contributions of the planned study. Pending acceptance, this study will be fully executed and the results reported thereafter.

2 Literature Review and Motivation

2.1 Three Key Papers

The works of González et al. (2022), Huang et al. (2020) and Sanford and Moosa (2015) critically underpin this study. As illustrated in Figure 1, these three papers motivate and encircle the gap in FIs' ORM toolkits that this study explores.

Figure 1. Depiction of the study's novelty at the intersection of three sequentially related constructs and their key motivating papers.



Supporting previous works, González et al.'s (2022) recent study found that insurers with greater quality and degree of enterprise risk management (ERM) lead to higher profitability, lower risk of insolvency and greater financial stability. Huang et al. (2020) explore aspects of what effective risk management means. They note the importance of quantitative (rather than qualitative) risk measurement, demonstrating a non-parametric approach to estimating value-at-risk (VaR) more accurately measures risk than traditional approaches, improving the effectiveness of risk management hedging strategies in commodity markets. An earlier work by Sanford and Moosa (2015) was instrumental in progressing ORM toward a more quantified and proactive approach for FIs. However, the Bayesian network-based tool remains limited by its reliance on inputs from a sample of experts and the underlying reactive objective of capital estimation. In light of the relationships between quantitative risk management tools, effective risk management and firm value, the current study investigates the feasibility and value of quantitative CFA for operational risks, building on Sanford and Moosa (2015).

2.2 Broader Literature

2.2.1 Data-driven ORM for Financial Institutions

Applications of data analytics for ORM within the financial services (FS) sector are mostly applied to banks. This is consistent with the level of regulation for operational risks across FIs. For example, the Australian Prudential Regulation Authority's (APRA's) operational risk regulation is most stringent for authorised deposit-taking institutions (APS 114 and 115 (APRA, 2013a, 2013b)), in comparison to insurance (GPS and LPS 118 (APRA, 2013c, 2019)) and superannuation (SPS 114 (APRA, 2013d)) entities.

The applications of data analytics to ORM in FIs are varied but can broadly be categorised into five main themes, representing the different ways data analytics is used to manage operational risks – risk identification, causal factors, risk quantification, risk prediction and risk decision-making (Table 1) (Cornwell et al., 2022). All five themes seek to enhance the efficiency and effectiveness of ORM. However, understanding the factors contributing to risks and their causal pathways is fundamental to proactive ORM, as reflected in the *Principles for Sound Management of Operational Risk* (Basel Committee on Banking Supervision, 2021). 'Black box' machine learning algorithms that predict losses do not lend themselves well to cause-effect analyses. Interpretable tools that help reveal key drivers of predicted losses offer more valuable outputs to improve risk management effectiveness and thus create firm value (Braumann, 2018).

2.2.2 Operational Risk Causal Factors Analysis

Many traditional approaches for establishing the causality of operational risks are based

Table 1. Summary of literature applying data analytics to ORM within FS[2].

	Risk Identification	Causal Factors	Risk Quantification	Risk Prediction	Risk Decision-making
Percentage of Research[3]	16%	25%	25%	13%	9%
Analytics Objective	Descriptive	Diagnostic	Predictive Diagnostic	Predictive	Prescriptive
Risk Perspective	Micro	Micro	Macro Multi-risk	Micro	Multi-risk
Analytics Techniques	Process Mining, Clustering	Association Rule Mining, Decision Tree, Natural Language Processing	Traditional Distribution Fitting (Loss Distribution Approach), Bayesian Network	Decision Tree, Artificial Neural Networks	Deep Reinforcement Learning
Data Inputs	• Risk, Incident and Loss (e.g., timing, consequence, monetary loss) • Technical System (e.g., system access, database queries, transaction characteristics) • Organisational Structure (e.g., role, business line, firm size) • Social/People (e.g., age, qualification, performance) • Macro-environmental (e.g., regulation, economy, country)				

on observation and past experiences (Covello & Mumpower, 1985). Within FS, the most universally practiced forms of causal factor identification and analysis (Chapelle, 2018; Valis & Koucky, 2009) are:

- <u>interviews</u> with or ad hoc learnings shared by staff on the front line of operations and experienced senior managers;
- <u>process mapping</u> or task analysis, where causes are identified by mapping out the tasks and associated risks for activities conducted in a process;

---

[2] The summary is based on the findings from Cornwell et al.'s (2022) comprehensive review of the discipline. Risk perspective is a three-level factor (micro, multi-risk, macro) indicating "the number of operational risks considered and the level of detail in which a study views them" (p. 5). See Cornwell et al. (2022) for more detail.

[3] The percentages do not sum to 100% since the remaining 12% of references were academic review papers, and hence not categorised into one of the five research themes.

- investigations and reviews of past losses that occurred internally or externally to the organisation (e.g., among peer organisations) or near-misses;
- risk and control self-assessments (RCSAs); and
- root cause analysis (e.g., using a bowtie tool).

RCSAs are often the principal operational risk assessment tool for FIs, typically performed annually and updated following significant loss events. Through workshop-style discussions among risk personnel, RCSAs evaluate the inherent likelihood and impact of significant operational risks, as well as the effectiveness of the controls designed to mitigate them, ultimately providing a residual risk measure (Chapelle, 2018). While broader than simply understanding loss causes, CFA is considered in RCSAs.

A more specific methodology for systematic and comprehensive cause-effect analyses is bowtie analysis. For a given risk event (centre ‘knot’ of the bowtie), (1) direct and indirect causes leading to the risk event are noted (left-hand triangle of the bowtie), (2) direct and indirect consequences flowing from the risk event are noted (right-hand triangle of the bowtie), and (3) (a) preventative controls between the causes and event and (b) corrective controls between the event and consequences are mapped to mitigate the occurrence and severity of the risk event, respectively (Chapelle, 2018). This process neatly allows multiple levels of causal failures to be identified and can inform key risk indicators and actions to address deficient controls. Bowties, indeed most CFAs, are usually applied in silos (i.e., to a single incident, type of event or process). This, however, does not allow commonalities in cause-effect relationships across risks to be uncovered, nor contagion of risks, which otherwise could enable more efficient mitigation.

Thinking further about the current qualitative and manual CFA methods, a lack of reliability and repeatability comes to mind. More specifically, the weaknesses are:

- labour intensive and costly;

- backward-looking and <u>reactive</u>;
- <u>biased</u> by assessors' subjective judgements and experiences, such that individuals having deep and broad knowledge of the risk event and possible causes and consequences are critical (although rare);
- <u>inconsistency</u> in assessments as assessors change over time, restricting reliable analysis of risk trends;
- limited by <u>human processing capacity</u> (a structure of four variables (Halford et al., 2005)), such that complex interrelationships cannot be inferred beyond a small subset of factors, nor similarities drawn from more than a few experiences; and
- <u>infrequent</u> and static snapshots from which organisations are making decisions, despite variation in the dynamic system between assessments.

In the past decade, there have been various applications of diagnostic data analytics techniques to gain a more objective and population-based understanding of the factors that contribute to operational risk events on a repeatable basis. These include studies on causal factor identification, risk contagion and understanding causal factor dynamics. Table 2 summarises these core study themes and notes key limitations.

Reflecting on all data-driven CFA approaches, there are three key conclusions. First, few studies on FIs that adopt a fully data-driven approach incorporate operational data in addition to loss data, as in asset reliability management applications. Lien (2012), as an example, develops fuzzy decision trees using database access logs in an insurance company to classify an employee's access to customer data as legitimate or inappropriate, thereby identifying employee and query characteristics indicative of inappropriate access.

Second, the granularity of insights gained into the causal mechanisms is somewhat limited by most existing approaches measuring risk in categories (e.g., Lien (2012) use "risk" or "no risk"). By comparison, capital estimation models consider

Table 2. Summary of the literature relating to the data-driven causal factors analysis theme.

	Causal Factor Identification	**Risk Contagion or Propagation**	**Causal Factor Dynamics**
Description	Identify <u>what</u> operational factors typically lead to operational losses	Identify common <u>risk propagation pathways</u> (i.e., loss events as causal factors to other risks (Deng et al., 2019))	Investigate <u>how</u> factors influence operational loss events
Analytics Techniques	Unsupervised, exploratory analytics (e.g., text mining)	Various	Bayesian Network, Decision Tree
Data Inputs	Loss data (internal or external)	Loss data	Expert elicitation, Loss and operational data
Example Studies	(Bouveret, 2019; Neil et al., 2009; Wang et al., 2018)	(Chernobai et al., 2008, 2011; Gao & Wang, 2021)	(Lien et al., 2011; Pika et al., 2013; Sanford & Moosa, 2015)
Limitations	Dependent on causes people report in incident reports and investigations	No insight into underlying factors FIs can proactively control	Assessor bias due to expert elicited data (dependence structures and probabilities)

operational risk on a continuum by adopting an actuarial mindset and using approaches with probabilistic structures. It is well understood that risk is an ever-changing phenomenon which lies on a continuous spectrum (McNeil et al., 2015). Hence, measuring the influence of causal factors on the probability of a loss event would provide a more realistic and detailed interpretation of a dynamic operational risk profile.

Finally, previous studies are often limited to a single type of operational loss. Sanford and Moosa's (2015) expert elicitation-based approach, however, takes a broader systems theory perspective, modelling the probability of multiple operational losses (payment failures, exposure management failures and regulatory/tax/legal failures) with consideration to socio-technical factors in an Australian wholesale bank (information and communication technology, staff knowledge, skill and capability). Incorporating multiple loss event types, considering the dependence of factors across them and allowing for risk contagion is crucial since “interconnections between business lines or risk categories give rise to avalanche-like effects” and “interconnections between upstream and downstream risks can induce network instabilities” (Mittnik & Starobinskaya, 2010, p. 389).

Overall, while data-driven approaches to CFA are beneficial in overcoming limitations of traditional approaches, there is substantial opportunity for advancement by viewing the problem from a purely data-driven and holistic systems perspective.

2.3 Motivation

As introduced in Section 2.1, the relationships between quantitative risk management tools, risk management effectiveness and firm value motivate the need for an improved quantitative approach to CFA for operational risks. The following expands on this.

Although risk management is often considered a defensive and compliance-focussed practice, effective risk management tools create firm value. There is substantial evidence highlighting the financial benefits of a higher degree and quality of ERM implementation in FIs. At its core, effective risk management reduces losses, thus reducing the volatility of cash flows and, therefore, a firm’s probability of default and insolvency (Chapelle, 2018; Gleißner, 2019; González et al., 2022; Ko et al., 2019). Greater financial stability attracts lower costs of capital (Gleißner, 2019; Krause & Tse, 2016; Shad et al., 2022) and enables superior earnings performance, profitability and stock returns (EY, 2013; Gleißner, 2019; González et al., 2022; PwC, 2015). Further, academics (Kaplan & Mikes, 2016) and practitioners (Chapelle, 2018; McKinsey & Company, 2020) find an effective risk management function improves productivity and frees up operational capacity to focus on strategic priorities and riskier projects, in turn accelerating growth and increasing enterprise value.

To realise these value-enhancing benefits, the mechanisms underlying effective risk management must be understood. Braumann (2018) analyses the influence of the components of ERM on the effectiveness of risk management. Braumann shows four practices that strengthen risk awareness, which positively affects risk management effectiveness – (i) the increased use of quantitative risk management tools (supported by Huang et al. (2020)), (ii) formal risk management processes and frameworks to establish a risk-centric organisational environment, (iii) formal risk reporting and communication and (iv) processes that integrate risk into strategic decision-making.

Financial risk management is well-developed with rigorous and quantitative models for credit, liquidity and market risks (Bessis, 2015; Dionne, 2013). However, non-financial risks are currently managed with largely qualitative and judgement-based assessments (e.g., RCSAs) (Chapelle, 2018) of limited effectiveness (Aven, 2016; McKinsey & Company & Operational Riskdata eXchange Association, 2017). Furthermore, while recognised as critically important for effective ERM in FIs (Nocoń & Pyka, 2019), maintaining strong regulatory and economic capital unintentionally drives reactive, outcome-focussed management, rather than proactive prevention. Arguably greater emphasis needs to be placed on the underlying factors and conditions exacerbating risks and communicating these to personnel on the first line of defence to be integrated into proactive, risk-informed decision-making. Such a paradigm shift is increasingly being encouraged by regulatory authorities, as in the *Principles for the Sound Management of Operational Risk* (Basel Committee on Banking Supervision, 2021).

In response, the application of data analytics to ORM is progressing the development of more robust and quantitative ORM tools (Araz et al., 2020; Nateghi & Aven, 2021). However, there is a lack of interpretable data-driven cause-effect analyses that could help in identifying the key drivers behind a predicted loss, thereby offering strategically more valuable outputs and ultimately enhancing firm performance.

3 Hypothesis Development (Idea)

The gap in FIs' ORM toolkits surrounding quantitatively assessing how causal factors influence operational risk has largely been limited by the research design of studies. Many studies are based only on loss event data, with no consideration of the factors and their conditions for non-events (i.e., normal operating conditions). Augmenting operational data with incident data to identify patterns, correlations and abnormalities during normal operating conditions as compared to incident inducing conditions underlies the comparatively advanced data-driven CFA research of condition monitoring (e.g., Onoda et al., 2009) and accident causality (e.g., Milana et al., 2019) in asset intensive and safety critical sectors, including mining, utilities and aviation. This concept is transferrable to operational risks in FIs. Indeed, there is a call for research to investigate the application of similar approaches to FS to extend the sector's quantitative operational risk models beyond aggregate capital estimations to exploring the underlying factors causing operational risks to be exacerbated (Cornwell et al., 2022).

Integrating elements of FIs' capital estimation models, such as their continuous probabilistic risk measures, would advance traditional methodologies for assessing individual operational risks in FIs, like RCSAs, that are predominantly based on likelihood-severity risk matrices, resulting in 'bucketed' risk ratings (e.g., "high", "moderate", "low"). Categorical risk ratings limit effective ORM in terms of (a) meaningfully differentiating and prioritising risks of the same rating and (b) reliably knowing the direction a risk is trending over time. This disparity between condition monitoring-esque approaches in asset intensive industries and probabilistic risk assessment in FS capital estimation methodologies motivates our first hypothesis.

H1: Integrating a FI's loss and internal operational data to model the probability of an operational loss event provides a platform for quantitative causal factors analysis.

H1’s investigation into the feasibility of leveraging a FI’s internal operational data to effectively model the probability distribution of operational loss events and their causal factors provides a benchmark for future studies improving on data-driven CFA.

With growing evidence of risk contagion and the interdependencies of operational factors to multiple losses, it is important to consider operational risks and their causal factors in a holistic system. Yet most existing data-driven CFA approaches are siloed to a specific risk (e.g., fraud). This leads to our second hypothesis based on systems theory.

H2: Modelling multiple operational loss events, rather than in isolation (as in H1), will enhance quantitative operational risk causal factors analysis.

The amalgamation of the two hypotheses represents this study’s novel idea – a quantitative methodology integrating a FI’s loss and operational data to analyse how a collection of causal factors influences the probabilities of multiple operational risk events.

4 Empirical Design

This study designs a Bayesian network-based methodology for quantitative CFA to several operational risks. It is applied in a real-world Australian insurance company. Within insurance, there are a variety of operational risks, relating to: the actions of staff; the behaviours and characteristics of customers, stakeholders and vendors, as well as the general public external to the organisation; the processes conducted to serve each existing and potential policyholder across the value chain; the organisation’s internal systems and technology, as well as those of key stakeholders in the supply chain. Many of these are similar to the operational risks present in other FIs, such as banks.

4.1 Data

The loss events investigated are various incidents of non-compliance by advisors in the company’s call centre. They encompass failing to follow company service requirements, internal process and regulatory requirements when speaking to customers on the phone, relating, for example, to an advisor's sales techniques or the process of asking and recording customer information.

The incidents analysed in the study are identified through a monthly standard audit process, which involves auditing a stratified random sample of calls from each advisor and reporting and validating incidents as appropriate. This sampling approach aims to ensure the data set of incidents unbiasedly represents all advisors and calls and is not skewed toward higher risk advisors or types of calls. While further details of the audit methodology cannot be explained due to confidentiality, the authors with extensive statistical sampling knowledge reviewed the methodology thoroughly and its consistency of application reflected in the data to ensure the rates of incidents derived reasonably reflect the company's normal operations.

Three data sets were sourced directly from the insurance company to provide the necessary event-level loss data and operational data for the period from 1 January 2019 to 30 June 2021. The data sets are:

- incident data, providing a record of the incidents of non-compliance that occurred during the period, including the timing, type of non-compliance and audit the incident was identified through;
- audit data, providing the complete history of standard monthly audits of calls conducted in the period, including the advisor behaviours and processes audited and their compliance for each call; and
- staff data, providing attributes and employment information of the advisors audited.

Augmenting the three data sets (linked using the unique audit and employee keys) provides a single data set for modelling, consisting of all operational factors for non-compliant and compliant calls audited. For non-compliant calls, all relevant types of non-compliance are marked with their respective indicator variables (i.e., the augmented data set contains a binary indicator variable for each type of non-compliance, where 0 represents no incident and 1 indicates an incident). There are data on 31 operational factors, relating to organisational structure, social or people and technical system factors. The variety of the types of operational factors the data encompass enhances the generalisability of the study's investigation into the practical application of the approach. Additional variables generated from the data are also included (e.g., the number of previous incidents for an advisor or time elapsed since a previous incident). Thorough exploratory data analysis (EDA) and cleaning are conducted in consultation with the industry partner to resolve any data quality issues, for example accuracy or completeness.

The 2.5 year sample of data is a sufficient quantity for the proposed methodology – there are 3,227 incidents of non-compliance reported across 13,562 audits. The frequency of incidents relating to different types of non-compliance is reviewed through EDA, and only types of non-compliance with sufficient frequency of occurrence are modelled. This seeks to avoid issues commonly reported as limitations of solely data-driven approaches to risk analysis, encompassing lack of data, extreme skewness and the zero-frequency problem. Furthermore, discussions with the insurance company confirmed that their business processes and structure were consistent throughout the entire sample period.

4.2 Tools

Bayesian networks (BNs) form the basis of our methodology. While a familiar tool in data-driven ORM (Table 1), the novelty of BNs in this study is in their application – in terms of the BN's intended use (i.e., operational risk CFA in FS) and how the ORM BN is learnt (i.e., solely from raw organisational data, not defined by experts). BNs for CFA are constructed to provide a granular, bottom-up representation of the operational risk environment, such that each operational factor and incident is modelled as a separate node. In contrast, applications of BNs for ORM in FS have predominantly been for capital estimation (e.g., Neil et al., 2005) with most emulating the traditional actuarial loss distribution approach – nodes for loss event frequency and severity, contributing to total loss, thus providing an aggregate measure of risk (Lambrigger et al., 2007; Mittnik & Starobinskaya, 2010). Employing a bottom-up perspective, similar to research managing safety incidents and asset reliability in nuclear energy (e.g., Pence et al., 2020), as well as medical diagnosis (e.g., Heckerman & Nathwani, 1992), enables causal relationships to be identified between tangible components of a FI's operations that they can manipulate. It not only quantitatively measures operational risk but also offers more valuable insights on what the drivers are and thus the actions that could help reduce risk. A BN-based approach also enables the investigation of the two hypotheses. BNs can combine multiple leading loss indicators (i.e., operating factors) with multiple outcome measures (i.e., losses and near-misses) and fundamentally, they are a probabilistic modelling methodology, associating each condition and outcome with a probability.

4.2.1 Bayesian Network CFA

A BN model for CFA is learnt using the insurance company's augmented data set following a two-phase methodology, consisting of data pre-processing and BN model fitting. All manipulation of data, statistical analysis and modelling is performed in R, a free and open-source programming language.

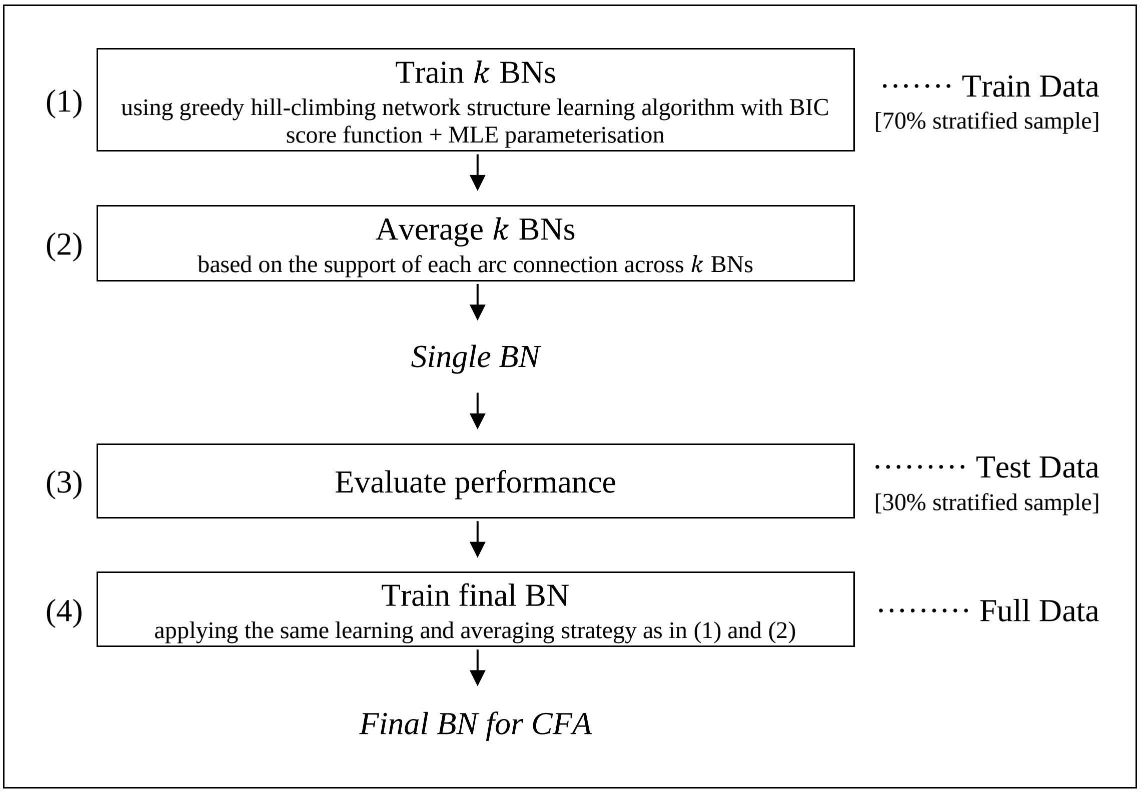
The pre-processing procedure comprises variable selection, discretisation and scaling to ensure the BN fits the data well and accurately reflects the insurer's operational risk environment. To avoid overfitting and computationally intractable dense networks (Bensi et al., 2011; Koller & Friedman, 2009), important features are selected prior to fitting the BN. We adopt Breiman's (2003) original clustering method, later refined by

Shi and Horvath (2006), which is a commonly applied and robust unsupervised feature selection algorithm based on the variable importance of random forests using the *randomForestSRC* package in R (Ishwaran & Kogalur, 2022). The majority of the operational factor variables in the data set are discrete, yet some are continuous. Given BN implementations are well-developed and parsimonious for discrete data, the continuous variables are discretised, enabling the specification of a discrete BN.

To fit the BN model, we adopt a learning strategy comprising the score-based greedy hill-climbing network structure learning algorithm using the BIC score function and MLE parametrisation, implemented using the *bnlearn* R package (Scutari, 2010). This is a theoretically robust learning strategy and is widely applied across the literature with reports of efficient and accurate results compared to alternative learning strategies (Hastie et al., 2009; Koller & Friedman, 2009; Kratzer et al., 2019; Scutari et al., 2019). Consistent with risk management theory (Rasmussen, 1997) and other analyses (e.g., bowtie (Chapelle, 2018)), we assume that operational factors can influence incidents but cannot be directly causally influenced by incidents. To reflect this in the network structure, arcs directed from incident nodes to operational factor nodes are restricted from the learning process by specifying them on a 'blocklist' of arcs.

In implementing the learning strategy, a train-average-test model fitting process is followed (Figure 2). A 70-30 train-test stratified partitioning protocol is used – a standard data science principle to avoid over or underfitting (Chollet, 2018; Helman et al., 2004). Stratification ensures both the training and testing subsets contain a representative sample of each type of non-compliance incident, which is necessary for imbalanced data as in this context. Model averaging, as opposed to model selection, is used, as research has shown it leads "to models that generalise well to *new* data" and yield accurate predictions (Heckerman, 2008, p. 21). The well-established methodology of BN

Figure 2. Bayesian network model fitting process.



model averaging using non-parametric bootstrapping is applied (Friedman et al., 1999; Nagarajan et al., 2013). A minimum of 100 BNs are averaged as per Friedman et al. (1999), although once the computational complexity is determined for the given data set, a larger number of networks will aim to be averaged (i.e., $k \geq 100$).

Prediction accuracy is a common assessment of the performance of BNs (Pérez-Bernabé et al., 2020; Wang & Li, 2021), so to provide an unbiased evaluation of the methodology's performance, the BN's accuracy on the test subset is measured by overall accuracy, recall, specificity and the geometric mean (G-mean) (Table 3). Recall and G-mean are particularly useful metrics in assessing class-imbalanced data sets, and hence are valuable in this context where incidents of non-compliance typically occur less often than not. The compute time to train the final BN (step (4)) is also reported.

Table 3. Measures of prediction accuracy.

Prediction Accuracy Measure	Description	Calculation	Objective
Overall Accuracy	Correctly classified instances	$\frac{TP + TN}{TP + FP + FN + TN}$	Maximise
Recall (or Sensitivity)	Class of interest classified correctly (positive accuracy)	$\frac{TP}{TP + FN}$	Maximise
Specificity	Negative accuracy	$\frac{TN}{FP + TN}$	Maximise
G-Mean	Ability to balance accurate classification of both the majority and minority classes	$\sqrt{Recall \times Specificity}$	=1

where TP, FP, FN and TN correspond to the accuracy of prediction of instances, as per the below confusion matrix

		Predicted	
		Loss	No Loss
Actual	Loss	True Positive (TP)	False Negative (FN)
	No Loss	False Positive (FP)	True Negative (TN)

4.2.2 Evaluation of Hypotheses

Hypothesis 1 is evaluated in two parts. The first part involves evaluating if, by applying the proposed BN methodology, a reasonable causal model that measures risk as a continuous probability can feasibly be built for a single operational loss event from historical incident and operational data. The model fitting process outlined above is conducted separately on several types of incidents of non-compliance using data on all operational factors and the single binary indicator variable for the incident type of interest, resulting in a BN for each incident type investigated. The reasonableness of the BNs is assessed based on the prediction accuracy metrics and by interpreting the network structures that indicate the flows of influence identified. The second part is evaluating if and how BN models can be used for CFA. From the BNs trained for each type of incident of non-compliance investigated, analysis and inference are conducted to gain insights into what and how factors influence the probability of non-compliance. A suite of inference queries and analyses on BNs that are meaningful for operational risk CFA are derived.

These are explained in the context of operational risk CFA and some demonstrated.

Hypothesis 2 is evaluated by comparing the single-risk BNs generated in H1 with a single BN modelling all the incident types simultaneously (i.e., a holistic systems perspective). The multi-risk BN is fit by applying the same model fitting process but to data on all operational factors and all incident type indicator variables, such that the resulting BN contains a node for each type of non-compliance investigated, in addition to relevant operational factors. Some of the operational risk CFA inference and analyses derived in H1 are performed on the multi-risk BN and the results compared to those in the single-risk BNs. This evaluates if and how the flows of influence change by taking a systems approach, or indeed what efficiencies in understanding the effect of causal factors on loss events can be gained. Evaluating this, as well as the prediction accuracy and computation time for the single-risk BNs compared to the multi-risk BN, inform if a systems approach enhances the quantitative operational risk CFA as hypothesised.

5 Conclusion

5.1 What's New?

There are three core novelties to this study's approach to data-driven CFA for FIs, representing new elements to the field's scholarly research and importantly ORM practice in industry. First, it integrates loss and operational data, akin to condition monitoring in asset intensive industries, to enable more objective and population-based inference, as well as more frequent assessments and monitoring, as compared to existing judgement-based and periodic analyses. Second, it measures operational risks as continuous probabilities with respect to changes in causal factors, rather than categorical risk ratings, which will allow more reliable prioritisation of risks and visibility on their trajectory. Third, it analyses the causal risk environment in greater depth to understand not only what causal factors, but also how they influence losses.

5.2 So What?

There are many practical implications of the quantitative approach to CFA proposed. Details on the stakeholders of relevance, their engagement with the method and the expected impact are outlined in Appendix A2. Most notably, FIs (banks, insurance and superannuation companies) will benefit, although regulators and risk consulting firms are also important stakeholders in guiding and enhancing ORM practice. The detailed insights about a FI's operational risk environment that the approach offers will inform more proactive and targeted mitigation strategies for risk managers, in turn ensuring efficient resource allocation and effective control and monitoring design. These benefits seek to drive more effective risk management within FIs, reducing the occurrence of financially, physically and reputationally costly losses. Ultimately, a relatively smaller VaR enhances firm stability, increases financial performance and improves operational efficiency – a mechanism for firm value enhancement.

5.3 Contribution

Overall, the study contributes a quantitative approach to assess and monitor how causal factors influence the probability of operational loss events in FIs. The BN-based approach constructs a model that captures the conditional probability distributions of and interrelationships between multiple operational risks and relevant operational factors. An empirical investigation of the approach on incidents of non-compliance by advisors in an Australian insurance company's call centre aims to demonstrate (H1) if and how a FI's historical loss and internal operational data can be leveraged for quantitative CFA and (H2) the information and efficiency gains of modelling operational risk environments from a holistic systems perspective. Not only will this study (an empirical causal factors study as classified in Cornwell et al.'s (2022) recently published framework) extend academic research on data-driven CFA, but it will enhance FIs' abilities to advance from reactive to proactive ORM, improving firm stability and financial performance.

5.4 Other Considerations

To achieve the practical outcomes of this research, a formal research collaboration with a leading industry partner for external risk management expertise and funding has been obtained, as well as a data sharing agreement with an Australian insurance company to provide data. All associated ethical clearance has been approved. The scope is achievable, although verging on too broad, so refinement of the types of non-compliance incidents and the causal factors modelled will be considered. Overall, the study is deemed low risk since the data has been secured, the tools are freely accessible, and the research team has appropriate expertise. The uniqueness and real-world nature of the data set and problem reduces competitor and obsolescence risk, positioning our study as highly impactful.

Funding

This work was supported by KPMG Australia and an Australian Government Research Training Program Scholarship.

Glossary

Term	Abbreviation (if used)	Definition
Bayesian network	BN	Within the family of probabilistic graphical models, BNs are directed acyclic graphs that concisely capture the conditional probabilistic dependence structure between a set of random variables, represented as nodes (Nagarajan et al., 2013).
Causal factors analysis	CFA	Analysis conducted in the ORM assessment process to identify the causes (direct, indirect or root causes) of incidents (Chapelle, 2018).
Enterprise risk management	ERM	"A systematic and integrated approach to the management of the total risks [operational, financial, strategic and external risks] that a company faces" (Dickinson, 2001, p. 360), in comparison to traditional siloed approach to risk management.
Exploratory data analysis	EDA	The process of conducting initial analysis and investigations on a data set to gain a thorough understanding of patterns and anomalies, as well as checking assumptions, before formal analysis and modelling. It often involves summary statistics and graphical visualisations.
Financial institution	FI	An organisation relating to the service of financial products or advice, including authorised deposit-taking institutions (or banks), insurance companies and superannuation entities.
Financial services	FS	The sector of industries relating to FIs (i.e., banking, insurance and superannuation).
Incident data		Records of event-level incidents (also referred to as loss or risk events) that occurred during core day-to-day operations in an organisation, including information about their timing, type and consequences.
Operational data		The raw data generated and collected from day-to-day business activities and processes in an organisation.
Operational risk		"The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events" (Basel Committee on Banking Supervision, 2006, p. 144), also referred to as non-financial risk. They include but are not limited to legal, regulatory, compliance, conduct, technology, data, reputational and change management risks (APRA, 2022).
Operational risk management	ORM	A continual recurring process involving identification, assessment, mitigation, monitoring, communication and reporting of operational risks to avoid the occurrence of incidents or near-misses (Chapelle, 2018).

Term	Abbreviation (if used)	Definition
Risk and control self-assessments	RCSA	The principal operational risk assessment tool for FIs to evaluate the likelihood and impact of operational risks, as well as assess control effectiveness (Chapelle, 2018).
Risk matrix		A two-dimensional matrix (typically 4-6 square) used to rank the likelihood of risks occurring on one axis and the severity of their consequences on the other, leading to an overall risk rating at their intersection.

References

APRA. (2013a). Prudential Standard APS 114 Capital Adequacy: Standardised Approach to Operational Risk: Australian Prudential Regulation Authority.

APRA. (2013b). Prudential Standard APS 115 Capital Adequacy: Advanced Measurement Approaches to Operational Risk: Australian Prudential Regulation Authority.

APRA. (2013c). Prudential Standard LPS 118 Capital Adequacy: Operational Risk Charge: Australian Prudential Regulation Authority.

APRA. (2013d). Prudential Standard SPS 114 Operational Risk Financial Requirement: Australian Prudential Regulation Authority.

APRA. (2019). Prudential Standard GPS 118 Capital Adequacy: Operational Risk Charge: Australian Prudential Regulation Authority.

APRA. (2022). Prudential Standard CPS 230 Operational Risk Management: Australian Prudential Regulation Authority.

Araz, O. M., Choi, T. M., Olson, D. L., & Salman, F. S. (2020). Role of Analytics for Operational Risk Management in the Era of Big Data. *Decision Sciences*, *51*(6), 1320-1346. https://doi.org/10.1111/deci.12451

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, *253*(1), 1-13. https://doi.org/https://doi.org/10.1016/j.ejor.2015.12.023

Basel Committee on Banking Supervision. (2006). International convergence of capital measurement and capital standards: A revised framework, comprehensive version: Bank for International Settlements.

Basel Committee on Banking Supervision. (2021). Revisions to the Principles for the Sound Management of Operational Risk: Bank for International Settlements.

Bensi, M., Der Kiureghian, A., & Straub, D. (2011). Bayesian network modeling of correlated random variables drawn from a Gaussian random field. *Structural Safety*, *33*(6), 317-332. https://doi.org/https://doi.org/10.1016/j.strusafe.2011.05.001

Bessis, J. (2015). *Risk Management in Banking* (4th ed.). Wiley. https://www.wiley.com/en-us/Risk+Management+in+Banking%2C+4th+Edition-p-9781118660218

Bouveret, A. (2019). Estimation of Losses Due to Cyber Risk for Financial Institutions. *Journal of Operational Risk*, *14*(2), 1-20. https://doi.org/10.21314/JOP.2019.224

Braumann, E. C. (2018). Analyzing the Role of Risk Awareness in Enterprise Risk Management. *Journal of management accounting research*, *30*(2), 241-268. https://doi.org/10.2308/jmar-52084

Breiman, L. (2003). Manual on setting up, using and understanding random forest, V4.0. https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Chapelle, A. (2018). *Operational risk management: best practices in the financial services industry*. John Wiley & Sons, Incorporated. https://ebookcentral.proquest.com/lib/bond/reader.action?docID=5613476

Chernobai, A., Jorion, P., & Yu, F. (2008). *The Determinants of Operational Losses*. https://www.fdic.gov/analysis/cfr/2008/april/chernobai-jorion-yu.pdf

Chernobai, A., Jorion, P., & Yu, F. (2011). The Determinants of Operational Risk in U.S. Financial Institutions. *Journal of Financial and Quantitative Analysis*, *46*(6), 1683-1725. https://doi.org/10.1017/S0022109011000500

Chollet, F. (2018). *Deep Learning with Python*. Manning Publications Co.

Cornwell, N., Bilson, C., Gepp, A., Stern, S., & Vanstone, B. J. (2022). The role of data analytics within operational risk management: A systematic review from the financial services and energy sectors. *Journal of the Operational Research Society*, 1-29. https://doi.org/10.1080/01605682.2022.2041373

Covello, V. T., & Mumpower, J. (1985). Risk Analysis and Risk Management: An Historical Perspective. *Risk Analysis*, *5*(2), 103-120. https://doi.org/10.1111/j.1539-6924.1985.tb00159.x

Deng, X., Yang, X., Zhang, Y., Li, Y., & Lu, Z. (2019). Risk propagation mechanisms and risk management strategies for a sustainable perishable products supply chain. *Computers & Industrial Engineering*, *135*, 1175-1187. https://doi.org/https://doi.org/10.1016/j.cie.2019.01.014

Dionne, G. (2013). Risk Management: History, Definition, and Critique. *Risk Management and Insurance Review*, *16*(2), 147-166. https://doi.org/10.1111/rmir.12016

EY. (2013). *Turning risk into results: How leading companies use risk management to fuel better performance*. https://web.actuaries.ie/sites/default/files/erm-resources/turning_risk_into_results_au1082_1_feb_2012.pdf

Faff, R. W. (2015). A simple template for pitching research. *Accounting & Finance*, *55*(2), 311-336. https://doi.org/https://doi.org/10.1111/acfi.12116

Faff, R. W., & Kastelle, T. (2016). Pitching Research for Engagement and Impact. http://ssrn.com/abstract=2813096

Friedman, N., Goldszmidt, M., & Wyner, A. (1999). *Data analysis with bayesian networks: a bootstrap approach*. Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, Stockholm, Sweden.

Gao, X., & Wang, Z. (2021). On modeling contagion in the formation of operational risk loss. *Journal of Operational Risk*, *16*(2), 1-17. https://doi.org/10.21314/JOP.2021.003

Gleißner, W. (2019). Cost of capital and probability of default in value-based risk management. *Management Research Review*, *42*(11), 1243-1258. https://doi.org/10.1108/MRR-11-2018-0456

González, L. O., Santomil, P. D., & Hoyt, R. E. (2022). The impact of ERM on insurer performance under the Solvency II regulatory framework. *The European Journal of Finance*, 1-25. https://doi.org/10.1080/1351847X.2022.2053180

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How Many Variables Can Humans Process? *Psychological Science*, *16*(1), 70-76. https://doi.org/10.1111/j.0956-7976.2005.00782.x

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Second ed.) [Textbook]. Springer.

Heckerman, D. (2008). A Tutorial on Learning with Bayesian Networks. In D. E. Holmes & L. C. Jain (Eds.), *Innovations in Bayesian Networks: Theory and Applications* (pp. 33-82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85066-3_3

Heckerman, D. E., & Nathwani, B. N. (1992). An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research*, *25*(1), 56-74. https://doi.org/https://doi.org/10.1016/0010-4809(92)90035-9

Helman, P., Veroff, R., Atlas, S. R., & Willman, C. (2004). A Bayesian network classification methodology for gene expression data. *Journal of Computational Biology*, *11*(4), 581-615. https://doi.org/10.1089/cmb.2004.11.581

Huang, J., Ding, A., Li, Y., & Lu, D. (2020). Increasing the risk management effectiveness from higher accuracy: A novel non-parametric method. *Pacific-Basin Finance Journal*, *62*, 101373. https://doi.org/https://doi.org/10.1016/j.pacfin.2020.101373

Ishwaran, H., & Kogalur, U. B. (2022). Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). *R Package*, *version 3.1.0*. Retrieved 1 July 2022, from https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf

Kaplan, R. S., & Mikes, A. (2016). Risk Management—the Revealing Hand. *Journal of Applied Corporate Finance*, *28*(1), 8-18. https://doi.org/https://doi.org/10.1111/jacf.12155

Ko, C., Lee, P., & Anandarajan, A. (2019). The impact of operational risk incidents and moderating influence of corporate governance on credit risk and firm performance. *International Journal of Accounting & Information Management*, *27*, 00-00. https://doi.org/10.1108/IJAIM-05-2017-0070

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press.

KPMG US, & The Risk Management Association. (2018). *Operational Risk Management Excellence Report (Executive Report)*. https://advisory.kpmg.us/articles/2018/2018-operational-risk-management-excellence-survey-report.html

Kratzer, G., Lewis, F. I., Comin, A., Pittavino, M., & Furrer, R. (2019). Additive Bayesian network modelling with the R package Abn. *arXiv preprint*.

Krause, T. A., & Tse, Y. (2016). Risk management and firm value: recent theory and evidence. *International Journal of Accounting and Information Management*, *24*(1), 56-81. https://doi.org/https://doi.org/10.1108/IJAIM-05-2015-0027

Lambrigger, D. D., Shevhenko, P. V., & Wüthrich, M. V. (2007). The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions. *Journal of Operational Risk*, *2*(3), 3-27.

Lien, C. C. (2012). The Application of Crisp and Fuzzy Decision Trees to Monitor Insurance Customer Database. *Information-an International Interdisciplinary Journal*, *15*(9), 3871-3876. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865035880&partnerID=40&md5=48bcdcc223c12dd51d2f971c2e641376

Lien, C. C., Ho, C. C., & Tsai, Y. M. (2011). *Applying fuzzy decision tree to infer abnormal accessing of insurance customer data*. 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011, Jointly with the 2011 7th International Conference on Natural Computation, ICNC'11, Shanghai. https://www.scopus.com/inward/record.uri?eid=2-s2.0-80053419517&doi=10.1109%2fFSKD.2011.6019676&partnerID=40&md5=7e390dadf2d761a6e5433162781125e0

McKinsey & Company. (2020). *The future of operational-risk management in financial services*. https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/the-future-of-operational-risk-management-in-financial-services

McKinsey & Company, & Operational Riskdata eXchange Association. (2017). *The future of operational risk*. https://managingrisktogether.orx.org/research/future-operational-risk

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools* (Revised ed.). Princeton University Press. https://books.google.com.au/books?id=l2yYDwAAQBAJ

Milana, D., Darena, M. S., Bettio, N., Cerruti, C., Siliprandi, G., Fidanzi, A., Cerioli, P., Silvestri, G., Tarasconi, F., Caserio, M., Botros, M., & Gabrielli, M. L. (2019). *Natural language understanding for safety and risk management in oil and gas plants*. Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079510865&partnerID=40&md5=cb2efd9719e5d9306098a7d2b4569354

Mittnik, S., & Starobinskaya, I. (2010). Modeling Dependencies in Operational Risk with Hybrid Bayesian Networks. *Methodology and Computing in Applied Probability*, *12*(3), 379-390. https://doi.org/10.1007/s11009-007-9066-y

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian Networks in R with Applications in Systems Biology* (First ed.). Springer. https://doi.org/https://doi.org/10.1007/978-1-4614-6446-4

Nateghi, R., & Aven, T. (2021). Risk Analysis in the Age of Big Data: The Promises and Pitfalls. *Risk Analysis*, *41*(10), 1751-1758. https://doi.org/10.1111/risa.13682

Neil, M., Fenton, N., & Tailor, M. (2005). Using Bayesian Networks to Model Expected and Unexpected Operational Losses. *Risk Analysis*, *25*(4), 963-972. https://doi.org/10.1111/j.1539-6924.2005.00641.x

Neil, M., Häger, D., & Andersen, L. B. (2009). Modeling operational risk in financial institutions using hybrid dynamic Bayesian networks. *Journal of Operational Risk*, *4*(1), 3-33. https://search.proquest.com/docview/223580181?accountid=26503

Nocoń, A., & Pyka, I. (2019). Sectoral analysis of the effectiveness of bank risk capital in the Visegrad Group countries. *Journal of Business Economics and Management*, *20*, 424-445. https://doi.org/10.3846/jbem.2019.9606

Onoda, T., Ito, N., & Yamasaki, H. (2009). *Interactive trouble condition sign discovery for hydroelectric power plants*. 15th International Conference on Neuro-Information Processing, ICONIP 2008, Auckland. https://www.scopus.com/inward/record.uri?eid=2-s2.0-70349145393&doi=10.1007%2f978-3-642-03040-6_81&partnerID=40&md5=b237c7491d46f4e46153f9d5bdf690e1

Operational Riskdata eXchange Association. (2022). *Annual Banking Loss Report*. Operational Riskdata eXchange Association. https://managingrisktogether.orx.org/loss-data/annual-banking-loss-report

Pence, J., Farshadmanesh, P., Kim, J., Blake, C., & Mohaghegh, Z. (2020). Data-theoretic approach for socio-technical risk analysis: Text mining licensee event reports of U.S. nuclear power plants. *Safety Science*, *124*. https://doi.org/10.1016/j.ssci.2019.104574

Pérez-Bernabé, I., Maldonado, A., Salmerón, A., & Thomas, D. (2020). MoTBFs: An R Package for Learning Hybrid Bayesian Networks Using Mixtures of Truncated Basis Functions. *The R Journal*, *12*, 321. https://doi.org/10.32614/RJ-2021-019

Peters, G. W., Clark, G., Thirlwell, J., & Kulwal, M. (2018). Global perspectives on operational risk management and practice: a survey by the Institute of Operational Risk (IOR) and the Center for Financial Professionals (CeFPro). *Journal of Operational Risk*, *13*(4), 47-88. https://doi.org/10.21314/JOP.2018.215

Pika, A., van der Aalst, W. M. P., Fidge, C. J., ter Hofstede, A. H. M., & Wynn, M. T. (2013). *Profiling event logs to configure risk indicators for process delays*. 25th International Conference on Advanced Information Systems Engineering, CAiSE 2013, Valencia. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84879878139&doi=10.1007%2f978-3-642-38709-8_30&partnerID=40&md5=56b1b19765cde97df72757b35d488b2e

PwC. (2015). *Risk in review: Decoding uncertainty, delivering value*. https://www.pwc.com/gr/en/publications/assets/pwc-risk-in-review-2015.pdf

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2-3), 183-213. https://doi.org/10.1016/S0925-7535(97)00052-0

Sanford, A., & Moosa, I. (2015). Operational risk modelling and organizational learning in structured finance operations: A Bayesian network approach. *Journal of the Operational Research Society*, *66*(1), 86-115. https://doi.org/10.1057/jors.2013.49

Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, *35*. https://doi.org/10.18637/jss.v035.i03

Scutari, M., Graafland, C. E., & Gutiérrez, J. M. (2019). Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. *International Journal of Approximate Reasoning*, *115*, 235-253. https://doi.org/https://doi.org/10.1016/j.ijar.2019.10.003

Shad, K., Lai, F.-W., Shamim, A., McShane, M., & Zahid, S. (2022). The Relationship between Enterprise Risk Management and Cost of Capital. *Asian Academy of Management Journal*. https://doi.org/10.21315/aamj2022.27.1.4

Shi, T., & Horvath, S. (2006). Unsupervised Learning With Random Forest Predictors. *Journal of Computational and Graphical Statistics*, *15*(1), 118-138. https://doi.org/10.1198/106186006X94072

Valis, D., & Koucky, M. (2009). Selected overview of risk assessment techniques. *Problemy eksploatacji*, 19-32.

Wang, Q., & Li, C. (2021). Evaluating risk propagation in renewable energy incidents using ontology-based Bayesian networks extracted from news reports. *International Journal of Green Energy*. https://doi.org/10.1080/15435075.2021.1992411

Wang, Y., Li, G., Li, J., & Zhu, X. (2018). *Comprehensive identification of operational risk factors based on textual risk disclosures*. 6th International Conference on Information Technology and Quantitative Management, ITQM 2018, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062043861&doi=10.1016%2fj.procs.2018.10.229&partnerID=40&md5=286d49a9ec755fa6a4411e93d3269477

Appendices

Appendix A: Research Pitches

The following tables are the approved research pitches submitted for Phase 2 of the PBFJ pre-registration publication submission process for this "Engagement & Impact" study *Value enhancement through improved data-driven operational risk management in financial institutions*. Appendix A1 presents the scholarly pitch of the research following Faff's (2015) Pitching Research Framework (PRF). Complementing this, Appendix A2 outlines the engagement and impact aspirations of this real-world research using Faff and Kastelle's (2016) template for pitching research for engagement and impact (PR4EI).

Appendix A1: Scholarly Pitch

<table>
<tr><td>Research Team</td><td>XXX</td><td>FoR Category</td><td>Operational Risk Management
[JEL: G20, G32, D81, C44]</td><td>Date Completed</td><td>23 May 2022</td></tr>
<tr><td>FOUR</td><td colspan="5">Four aspects of BIG picture framing</td></tr>
<tr><td>A. Working Title</td><td colspan="5">Value enhancement through improved data-driven operational risk management in financial institutions</td></tr>
<tr><td>B. Basic Research Question</td><td colspan="5">How can a financial institution effectively leverage internal operational and incident data to model how causal factors influence the probability of operational risk events, as a tool to improve the effectiveness of risk management, reduce risk exposure and thus enhance financial performance?</td></tr>
<tr><td>C. Key Paper(s)</td><td colspan="5">Sanford, A., & Moosa, I. (2015). Operational risk modelling and organizational learning in structured finance operations: A Bayesian network approach. Journal of the Operational Research Society, 66(1), 86-115. doi:10.1057/jors.2013.49

González, L. O., Santomil, P. D., & Hoyt, R. E. (2022). The impact of ERM on insurer performance under the Solvency II regulatory framework. The European Journal of Finance, 1-25, https://doi.org/10.1080/1351847X.2022.2053180

Huang, J., Ding, A., Li, Y., & Lu, D. (2020). Increasing the risk management effectiveness from higher accuracy: A novel non-parametric method. Pacific-Basin Finance Journal, 62, 101373. https://doi.org/https://doi.org/10.1016/j.pacfin.2020.101373</td></tr>
<tr><td>D. Motivation/Puzzle</td><td colspan="5">While risk management is often considered a defensive and compliance-focussed practice, effective risk management tools create firm value. Research has shown more effective risk management enhances firm stability (by reducing volatility and insolvency risk), increases financial performance (including profitability and stock returns) and improves operational efficiency (through increased operational and growth capacity).

Financial risk management is well-developed with rigorous and quantitative models for credit, liquidity and market risks. Non-financial, or operational, risks, however, are currently managed with largely qualitative and manual assessments. The effectiveness of these tools is limited by their periodic and static application, the subjective biases of individual assessors as well as their reactive and backward-looking nature.

Research has progressed to developing more robust and quantitative data-driven ORM tools. Using data analytics to identify what operational factors cause operational loss events constitutes a major aspect of these advancements. Yet, existing techniques do not allow for in-depth insights to be uncovered about how the causal factors are influencing the probability of a loss event. Such insights would provide decision-makers with near real-time information to act on, enhancing the effectiveness of ORM and thus driving enterprise value.</td></tr>
<tr><td>THREE</td><td colspan="5">Three core aspects of any empirical research project i.e. the “IDioTs” guide</td></tr>
<tr><td>E. Idea</td><td colspan="5">The core idea of this study is to augment raw operational data with incident and loss data to gain an improved understanding of the effect of causal factors on the probability of operational loss events. This will entail developing and evaluating, theoretically and empirically, a model that captures the conditional probability distributions of and interrelationships between multiple operational risks and various relevant operational factors. Several outputs about the mechanics of operational risk events in financial institutions are expected, including:
the collection of causal factors that influence the probability of a single or multiple incidents;
the relative importance of causal factors contributing to the probability of an occurrence;
the settings or thresholds of causal factors at which the probability of an incident changes substantially.</td></tr>
<tr><td>F. Data</td><td colspan="5">This study will be investigated on incidents of non-compliance by advisors within an Australian insurance company’s call centre. The data from a 2 to 3 year historical period is two-fold:
<u>event-level loss data</u> – a record of the occurrences of non-compliance, including timing and type of non-compliance; and
<u>operational data</u> – data from operational factors that may relate to the incidents of non-compliance, measured or observed at the time of the incident, for example the advisor’s characteristics, organisational structure factors and insurance product details.
The incident data will represent the response variable, informing the probability of the loss event, and the operational data will contain the independent variables, or causal factor inputs. The data sets will be augmented to generate a single data set for modelling, consisting of the observations of all operational factors for both non-compliant and compliant calls audited in the insurance call centre.</td></tr>
</table>

	An industry partner, who we have collaborated with for the research, are providing the data sets. Potential data quality issues, relating to accuracy or completeness, will be externally validated by the industry partner and carefully managed when modelling.
G. Tools	Determined from an extensive literature review, Bayesian networks will be applied, using the standard data science process for model development and performance evaluation. Several network models will be trained and parameterised using various learning algorithms and their predictive capability will be evaluated using relevant accuracy and performance metrics. The model's insights will also be reviewed in consultation with domain experts from the industry partner to consider the model's reflection of reality and practicality. Manipulation of data and the statistical analysis and modelling will be performed in R, a free and open-source programming language, using appropriate packages, such as *bnlearn*. Other software (e.g., Python) will be used where necessary.
TWO	**Two** key questions
H. What's New?	1. The core novelty of this research is that it integrates an organisation's history of incident data (currently collected for regulatory or second-line risk purposes) with internal operational data to enable data-driven operational risk causal factors analysis. This differs from most existing root cause analysis methodologies used in financial institutions, such as manual post-hoc incident investigations or periodic assessments relying on expert elicitation. The study seeks to apply a condition monitoring-esque approach as in asset-intensive industries to financial services. 2. This study extends existing data-driven causal factors research by representing operational risk as a continuous probability with respect to changes in causal factors, rather than as categorical or 'bucketed' risk ratings. 3. Additionally, the approach developed in this study is unique in the depth of understanding it offers into *how* causal factors influence operational risk events (i.e., collective and compound effects, relative importance and critical risk thresholds), compared to simply *what* factors, as in existing literature.
I. So What?	For each novelty in Item H (in order): 1. Leveraging these data sources in organisations enables more objective and population-based inference, as well as more frequent causal factor assessments and monitoring. 2. A continuous probabilistic measure of operational risk alleviates major limitations with discrete risk ratings surrounding prioritisation of multiple risks in a single risk rating 'bucket' and visibility on how risks are trending over time. 3. More detailed insights about an organisation's operational risk environment will inform more proactive and targeted mitigation strategies for risk managers, in turn ensuring efficient resource allocation and effective control and monitoring design. Overall, these benefits seek to drive more effective risk management within financial institutions, reducing the occurrence of financially, physically or reputationally costly losses. Ultimately, a relatively smaller value-at-risk reduces regulatory and economic capital requirements, mobilising capital for growth, in turn increasing profitability and enterprise value.
ONE	**One** bottom line
J. Contribution	This study practically applies a quantitative, data-driven approach to assess and monitor the causal factors influencing the probability of incidents of non-compliance by advisors in an Australian insurance company's call centre, from which precise and actionable control and mitigation efforts can be derived. Not only will this extend academic research on data-driven causal factors analysis, but it will enhance practitioners' abilities to advance from reactive to proactive ORM in financial institutions, in turn improving firm stability and financial performance.
K. Other Considerations	• A formal research collaboration with a leading industry partner for external risk management expertise and funding has been obtained, as well as a data sharing agreement with an Australian insurance company for the provision of data. • Ethical clearance has been approved. • The study is deemed low risk since the data has been secured, the tools required are freely accessible and the research team involved has appropriate expertise. The uniqueness and real-world nature of the data set and problem reduces competitor and obsolescence risk. • The scope is achievable, although verging on too broad, so refinement of the types of non-compliance incidents and the causal factors modelled will be considered.

Appendix A2: Engagement and Impact Pitch

Research Team	XXX	**FoR Category**	Operational Risk Management *[JEL: G20, G32, D81, C44]*	**Date Completed**	23 May 2022
FOUR	**Four** aspects of BIG picture framing				
A. Working Title	Leveraging data to advance operational risk management				
B. Basic Impact Goal	A data-driven approach to operational risk causal factors analysis will provide financial institutions with valuable insights about their interconnected risk environment to aid proactive decision-making and effective resource allocation, ultimately reducing the frequency and severity of financially, physically and reputationally costly losses.				
C. Key Industry / External Triggers	Peters, G. W., Clark, G., Thirlwell, J., & Kulwal, M. (2018). Global perspectives on operational risk management and practice: a survey by the Institute of Operational Risk (IOR) and the Center for Financial Professionals (CeFPro). *Journal of Operational Risk*, 13(4), 47-88. doi:10.21314/JOP.2018.215 KPMG US, & The Risk Management Association. (2018). *Operational Risk Management Excellence Report (Executive Report)*. Retrieved from https://advisory.kpmg.us/articles/2018/2018-operational-risk-management-excellence-survey-report.html Basel Committee on Banking Supervision. (2021). *Revisions to the Principles for the Sound Management of Operational Risk*. In: Bank for International Settlements.				
D. Motivation/Problem	In the increasingly complex, dynamic and regulated operating environments of financial institutions, practitioners recognise the value and importance of highly effective and proactive ORM tools in limiting the direct and indirect financial losses that can result from the large volumes and varieties of operational risks. Across the financial services sector, there is a growing expectation that operational risk managers do not simply comply with capital adequacy regulation, but that ORM is a critical revenue-protection and -driving mechanism. However, to achieve this, substantial innovation and transformation of existing subjective, static and costly ORM tools is required. Risk professionals acknowledge that big data analytics will allow more proactive, real-time and effective ORM. Yet, such holistic and flexible systems which industry demand have not been robustly developed, tested, nor implemented.				
THREE	**Three** core impact dimensions				
E. Stakeholders	• Financial Institutions – banks, insurance and superannuation companies are a key beneficiary of this research, since the implementation of a robust data-driven approach to understand what factors lead to operational losses would seek to improve decision-making and reduce the negative consequences of operational risk events. • Risk Consulting Firms – since risk consulting firms offer best practice advice, services and products relating to enhancing organisations' ORM frameworks, processes and policies, understanding and developing such a system for commercialisation to meet the demands of their clients is of importance. Risk consulting firms are beginning to invest in such developments to expand their services and capability, yet this is limited by their financial and human resources. In particular, there are knowledge and expertise gaps surrounding risk management practice in the context of advanced data analytics. However, risk consulting firms are uniquely positioned to develop a holistic data-driven framework given their access to multiple organisations across a range of industries for the provision of data. Hence, a partnership with a top-tier risk consulting firm in Australia has been established as part of this research. • Regulators – as recognised as a key external trigger, regulatory bodies of financial institutions stipulate operational risk measurement, capital and management requirements, with the primary intent to protect consumers and prevent market failure. Thus, understanding, encouraging and regulating effective strategies and best practice for ORM is important. It is critical that the approach developed aligns to current ORM regulation, otherwise regulators could block the solution. Additionally, the regulators could use the causal factor insights from the approach to inform regulatory settings. All stakeholders are interconnected and interact with one another reasonably cooperatively. However, within the group of risk consulting firms, firms are competitors, vying for specific organisations as clients for commercial gain. Regulators and risk consultants share similar objectives in				

	their interactions with financial institutions, yet regulators hold greater control over the financial institutions, while financial institutions have greater bargaining power over the risk consulting firms. Regulators' decisions also greatly influence the work of risk consultants, and this flow of influence also works conversely (although to a lesser extent), as consultants advise regulators on regulatory reform and current industry practice.
F. Value Proposition	• The financial institution providing data for this study will most directly benefit from the research. For this study, it will provide the Australian insurance company with detailed insights into the operational factors that influence the probability of incidents of non-compliance by advisors in their call centre. From this, targeted control and mitigation strategies could be implemented to reduce non-compliance and enhance operations in the future. • The partnering risk consulting firm will be a secondary beneficiary of the research, providing them with an example implementation of the application of data analytics to analysing operational causal risks. Such an approach could be commercialised into a product to assist financial institutions in proactively assessing and monitoring the causal factors of operational risks. Indeed, the theoretically robust solution based on academic rigour may position the firm as an industry leader.
G. Resources	To realise this impact, • early engagement with a financial institution is required for the provision of high quality and large quantities of detailed event-level loss and operational data; • collaboration with a risk consulting firm to gain access to risk consultants for external risk management expertise and insight into current industry practice would enhance the research; and • funding for research assistants and expenses associated with information gathering and dissemination is also required from the industry partners. These resources have been accessed through a partnership with a top-tier risk consulting firm in Australia, and a data sharing agreement with an Australian insurance company has been arranged.
TWO	**Two** key impact signals
H. Communication Strategy	Collaborative meetings and agreed deliverables through the partnerships established will be the primary mechanisms for communicating the research process and results to the financial institution who provides data and the risk consulting firm. These engagements will be collaborative in nature, such that industry partners will also contribute their domain expertise and knowledge. Subsequently, consulting, publications and social media by the partnered risk consulting firm, incentivised by their commercial interest, will continue to disseminate the work and achieve the target impact for financial institutions. Opportunities to publish industry or academic publications and present at conferences to commence dialogue with financial institutions and regulators will also be sought.
I. Metrics	Feedback from the financial institution involved in the research surrounding the direct impact of the data-driven approach on their operational risk environment, people and organisation will be a valuable measure of success, including metrics such as the number of incidents of non-compliance. In time, once the solution is commercialised, the partnership with the risk consulting firm will provide revenue and the number of organisations adopting the approach as metrics.
ONE	**One** bottom line
J. Impact	The output of this research will allow financial institutions to have a greater understanding of the operational loss drivers, indicators of changing risk profiles and complex interconnections in their risk environment. The implementation of the data-driven approach to organisations will provide a more efficient and informed ORM process resulting in fewer and less catastrophic losses.
K. Other Considerations	• Negotiations surrounding the ownership of intellectual property, as well as confidentiality concerns and the use of an embargo have been completed under the partnership and data sharing arrangements. • The data-driven approach may not be eagerly adopted by the stakeholders due to uncertainty associated with a lack of technical understanding or due to limitations in financial resources. Hence, an effective communication strategy will be critical. The industry partner is aware of these potential risks, however there is a strong belief that the industry will soon be characterised by a data-driven approach and hence, there is a clear opportunity to benefit from the first-mover advantage.