



Extracting Best-Practice Using Mixed-Methods

Insights and Recommendations from a Case Study in Insurance Claims Processing

Erik Poppe · Anastasiia Pika · Moe Thandar Wynn · Rebekah Eden · Robert Andrews · Arthur H. M. ter Hofstede

Received: 15 January 2020 / Accepted: 10 December 2020 / Published online: 20 April 2021
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2021

Abstract *Problem Definition:* Queensland’s Compulsory Third-Party (CTP) Insurance Scheme provides a mechanism for persons injured as a result of a motor vehicle accident to receive compensation. Managing CTP claims involves multiple stakeholders with potentially conflicting interests. It is therefore pertinent to investigate whether ‘best practice’ for claims processing can be identified and measured so all claimants receive fair and equitable treatment. The project set out to test the applicability of a mixed-method approach to identify ‘best-practice’ using qualitative, process mining, and data mining techniques in an insurance claims processing domain. *Relevance:* Existing approaches typically identify ‘best practice’ from literature or surveys of practitioners. The study provides insights into an alternative, mixed-method approach to deriving best practice from historical data and domain knowledge. *Methodology:* The study is a reflective analysis of insights gained from a practical application of a mixed-method approach to determine ‘best practice’. *Results:* The mixed-method approach has a number of benefits over traditional approaches in uncovering best practice process behavior from historical data in the real-world context (i.e., can identify process behavior differences between high and low performing cases). The study also highlights a number of challenges with regards to the quality and detail of data that needs to be available to perform the analysis. *Managerial Implications:* The ‘lessons learned’ from this study

will directly benefit others seeking to implement a data-driven approach to understand a ‘best-practice’ process in their own organization.

Keywords Process mining · Best-practice · Insurance claim processing · Case study · Mixed-method

1 Introduction

Best practice is formally defined as “a successful way to treat a particular problem that may need to be adapted in skillful ways in response to prevailing conditions” (Mansar and Reijers 2007). Simply put, a best practice is the process behavior that results in *successful* outcomes. According to Cho et al. (2017), a process is considered successful (i.e., a best practice) based on the extent it improves performance dimensions as present in the Devil’s Quadrangle (i.e., time, cost, quality, and flexibility) (Brand and Van der Kolk 1995; Reijers and Mansar 2005).

These performance dimensions, precursors to determining if a process is best practice, are highly generalizable, but have limited transferability to domains where the notion of performance is associated with more than just processing tasks as quickly or as cheaply as possible. Moreover, existing approaches to identify ‘best practice’ usually derive best practice from literature or surveys of practitioners rather than in a data-driven way (Reijers and Mansar 2005; Christmann 2000). As such, domain-specific insights based on data are largely absent from literature.

With the increasing volume of data recorded in IT systems, organizations are better able to manage and improve their processes. Rather than relying on highly generic best practices based on performance dimensions that may not be relevant to their organization,

Accepted after two revisions by Jörg Becker.

E. Poppe · A. Pika · M. T. Wynn · R. Eden · R. Andrews (✉) · A. H. M. ter Hofstede
Queensland University of Technology, 2 George Street,
Brisbane, Australia 4000
e-mail: r.andrews@qut.edu.au

organizations can perform a data driven comparison of multiple organizational units to identify their own best practice processes.

We argue, that to identify domain-specific best-practices, first requires an understanding of the performance dimensions relevant to the organization. The best practice can then be identified as the behavioral difference between cases attaining high scores on these performance dimensions and cases not achieving such scores (or achieving them to a lesser degree).

In this study, we investigate best practice in injury compensation claims processing, a complex domain where parties (e.g., claimants, insurers, law firms, and health providers) often have conflicting interests. Queensland's Compulsory Third-Party (CTP) Insurance Scheme provides a compensation mechanism for persons injured as a result of a motor vehicle accident. Despite legislation mandating certain milestones for claims processing, the QLD Motor Accident Insurance Commission (MAIC) sees significant variations in the processing and outcomes of CTP claims. These variations indicate best-practice guidelines may be needed to ensure consistent and fair outcomes. It is therefore pertinent to investigate whether 'best practice' for claims processing can be identified and measured so claimants receive fair and equitable treatment across all scheme participants. This research project, therefore, set out to test the applicability of a mixed-method approach to identify 'best-practice' using qualitative, process mining, and data mining techniques.

We adopt a novel mixed-method approach (Wynn et al. 2019) which leverages claims processing data from multiple insurers, and interview data from process experts. We first identify dimensions, variables and measures of best performance relevant to the domain and then identify best practice behavior using process data. This mixed-method approach helps overcome limitations associated with using only qualitative or quantitative methods, providing a rich understanding of the domain and objective insights into how the best practice process unfolds.

This paper extends upon (Wynn et al. 2019) through applying the mixed-method approach (summarized in Sect. 3) to a completed real-world case study. Section 2 presents related work while Sect. 4 discusses the case scenario in the domain of injury compensation claim processing. Reflecting on challenges faced, and lessons learned from the case study (Sect. 5) led to key contributions of this paper. Section 6 provides a synthesis of these insights which will benefit others in applying our mixed method approach when determining best practice in other domains.

2 Related Work

Best practices are proffered to be associated with improvements in process efficiency, effectiveness, and quality (Mansar and Reijers 2007; Cho et al. 2017). As such, they are of interest to numerous domains, including business process management and service delivery (Rondini et al. 2018). Existing work on best practices, has mostly used either quantitative as in (Kis et al. 2017) or qualitative as in (Mansar and Reijers 2007) techniques on their own.

On the qualitative side, best practice has been identified using literature reviews, interviews, and focus groups. In Reijers and Mansar (2005), the authors derive best practice by reviewing business process redesign literature. In Mansar and Reijers (2007), the authors extended this work using a descriptive survey to identify 10 best practices, mapping them to the Devil's Quadrangle of cost, quality, time, and flexibility. However, these studies identify best practices based purely on perceptual data. Objective data of how the process unfolds are not examined.

On the quantitative side, historical data has been used to identify deviations from best practice. However no data-driven approach for deriving best practices could be identified. For instance, in del Rio-Ortega et al. (2012), the authors developed a tool to quantitatively measure performance indicators specified by the IT department, but no comprehensive qualitative methods and analysis techniques were used. Others quantitatively assess the impacts of best practices, derived from literature, on organizational performance (Christmann 2000). Several generic process mining methodologies recognise the need for stakeholder input (De Weerd et al. 2013; van Eck et al. 2015), but these approaches focus on diagnosing processes (i.e., identifying problems) rather than deriving best practice through process mining. A related area of study is that of decision mining (Rozinat and van der Aalst 2006; De Leoni and van der Aalst 2013; Mannhardt et al. 2016; De Smedt et al. 2017) which aims at the detection of data dependencies that affect the routing of a case. These data dependencies are used to derive constructs such as decision or regression trees (also referred to as prediction trees) (De Leoni et al. 2016) that can be used to cluster cases with analysis then performed on each of the clusters. Such approaches are not specifically targeted at identifying best practice.

While each perspective provides insights into best practice, a more complete understanding can be developed using a mixed method design (Venkatesh et al. 2016). In process mining, quantitative techniques are limited to data contained within event logs. Surface level understanding of what the attributes represent and data quality issues that are often present in these data sets can lead to spurious

findings (Bose et al. 2013). A qualitative approach can help mitigate these limitations through (i) providing contextual understanding which is necessary when exploring domain specific settings (e.g., best practice claims management); (ii) providing deeper insights into representations underlying data sets (e.g., identifying best practice dimensions).

While mixed methods are clearly beneficial for establishing completeness, developing inferences, and corroborating findings, they are still rarely applied in process mining and the broader Information Systems fields (Venkatesh et al. 2016; van der Aalst and Dustdar 2012). This is also reflected within our specific context of injury compensation claims processing, where existing studies have built based on either qualitative or quantitative approaches. In Francis et al. (2009), the authors consider fundamental aspects of best practices for accident compensation claims management reviewing literature. Conversely, in Andrews et al. (2018), the authors used data mining and process mining techniques to identify key factors contributing to delays in claims processing, but did not explore how qualitative methods can complement their studies to identify best practice.

3 Approach

This section provides a brief summary of a three-stage approach to determine the best practice of a specific domain (see Wynn et al. (2019) for a complete description of the method). The approach (Fig. 1) is indicative of a largely sequential mixed methods design (Venkatesh et al. 2016), with qualitative data collected and analyzed prior to the analysis of objective quantitative data as outlined below.

In the first stage, a domain-specific understanding of ‘best-practice’ is developed using qualitative methods. Using semi-structured interviews, stakeholders from the domain who have been identified through purposeful and theoretical sampling (Newman et al. 2013; Glaser 1998) are engaged to determine key dimensions of ‘best-practice’. The interviews are inductively analyzed using techniques from grounded theory, which involved open coding to identify best practice themes, constant comparison to resolve the best practice themes into a meaningful set, and theoretical coding (Glaser 1998) to identify relationships between themes. The outcome of this stage is the identification of performance dimensions (e.g., process timeliness), which can be comprised of multiple measurable variables (e.g., legislative timeliness, overall timeliness). The refined set of best performance dimensions and variables form the foundation of the second stage.

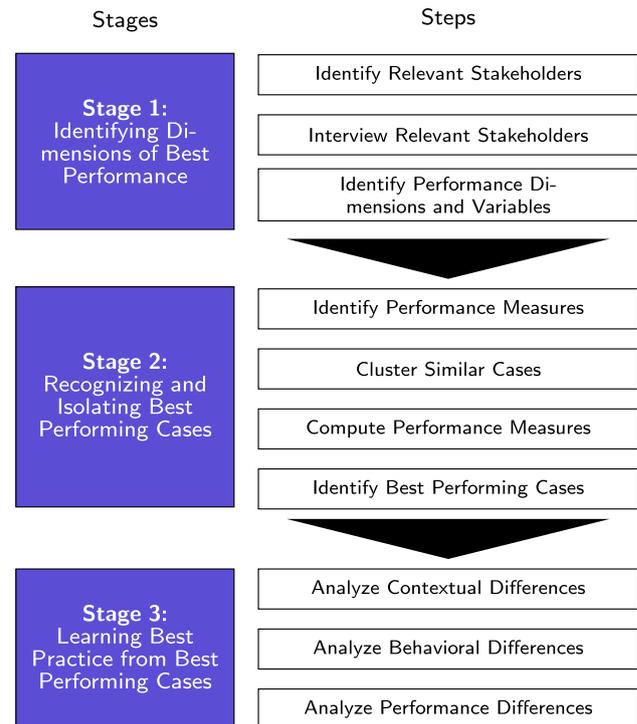


Fig. 1 A mixed-method approach to uncovering a best practice process

In the second stage, the best-performing cases are identified and isolated through analyzing objective historical data in the form of event logs. An event log is defined as a multi-set of traces (van der Aalst and Van Dongen 2013) comprised of the timestamps of the various activities involved in the process, as well as data related to the characteristics of both the case and case participants. To identify and isolate the best-performing cases involves four steps. Firstly, based on the performance dimensions and variables identified in *Stage 1*, the event logs are analyzed to identify appropriate measures for each performance variable. The measures can be directly observable or latent, whereby a latent variable is not concrete nor directly observable but can be inferred (MacKenzie et al. 2011). Subsequently, in the second step, cluster analysis is performed to minimize confounding (Greenland et al. 1999) by grouping claims together, based on case characteristics that are expected to behave in similar ways. In the third step, performance measures are calculated for each case, aggregated across performance dimensions and then combined into an overall performance value. Binning (i.e., discretization) (Aggarwal 2015) is then performed to label best and worst performing cases.

In the third stage, the best practice process behavior is identified through performing various types of data mining and process mining. Data mining is an umbrella term which refers to “the study of collecting, cleaning,

processing, analyzing, and gaining useful insights from data” (Aggarwal 2015). Process mining is a research discipline which “sits between machine learning and data mining on the one hand and process modeling and analysis on the other hand” and provides techniques for “extracting knowledge from event logs” (van der Aalst 2016). Specifically, the third stage involved analyzing contextual differences between high performing and low performing cases by applying supervised machine learning techniques (e.g., logistic regression, decision trees and random forests) (Aggarwal 2015); analyzing differences in process behavior by applying process discovery (Leemans et al. 2014) and process analysis (Mannhardt et al. 2015) approaches; and analyzing performance differences between high performing and low performing cases (Wynn et al. 2017).

While a largely sequential mixed-methods design was followed, during all three stages, qualitative member-checking strategies (Creswell and Miller 2000) have been used to validate key findings with stakeholders.

4 Case Scenario: Queensland CTP Claims Processing

The case study was performed in close collaboration with the Motor Accident Insurance Commission (MAIC) in Queensland¹ and the Nominal Defendant. In Queensland, the injury compensation claims scheme, known as the Compulsory Third Party (CTP) scheme, is governed by the Motor Accident Insurance Act 1994 and is underwritten by four licensed, commercial insurers who accept applications for insurance and manage claims on behalf of policyholders. The scheme is overseen by MAIC with the Nominal Defendant (ND), a statutory body, responsible for managing claims where the ‘at fault’ vehicle is unregistered or unidentified. Despite legislation mandating certain milestones for claims processing and providing various pathways for claims to be progressed and finalized, MAIC sees significant behavioral and performance variations in CTP claims processing and variations in costs and duration of claims. These variations raised the question whether best practice process behavior can be identified to ensure consistent and fair outcomes for all participants.

The case study explored the following three inter-related research questions:

- Q1 How can best practice in CTP insurance claim processing be *defined*?
- Q2 How can best practice in CTP insurance claim processing be *measured* given the data that is collected by scheme participants?

- Q3 How can best practice *process behavior* in CTP insurance claim processing be *identified from data*?

We adopted the mixed-method approach (described in Sect. 3 and detailed in (Wynn et al. 2019)) drawing on qualitative and objective quantitative data to uncover a best practice CTP process (Q1–3) and conducted a detailed study to demonstrate how such behavior could be mined from process and case data.

In terms of qualitative data, semi-structured interviews were conducted and analyzed using techniques from grounded theory (Glaser 1998). In terms of the quantitative aspects of our approach, we first focused on historical claims data collected by MAIC from all scheme participants (four commercial insurers plus ND). We conducted further analysis using ND process data as an exemplar case. We were able to use data values of several attributes known to affect claim performance to cluster cases. Case performance was derived from the clusters and process models derived based on the comparative analysis of ‘high performing’ and ‘low performing’ cohorts.

5 Case Study Findings

In the following sub sections, we reflect on our experience in conducting each stage of the approach in the case study (Stage 1: Identifying dimensions of best performance; Stage 2: Recognizing and isolating best performing cases; Stage 3: Learning best practice from best performing cases). For each stage, we present our findings separately in four sub-areas: (1) objective and methods used, (2) challenges encountered, (3) results, and (4) lessons learned.

5.1 Stage 1: Identifying Dimensions of Best Performance

5.1.1 Objective and Methods Used

We completed Stage 1 of the proposed approach, which involved conducting semi-structured interviews (Myers and Newman 2007) with multiple stakeholders to form a contextual understanding of a CTP ‘best practice’ process, with the goal of: (1) unpacking key dimensions underlying best performing cases; and (2) identifying potential variables, observed or latent, that can be used to assess the dimensions.

To ensure multiple perspectives were captured, we first performed preliminary meetings with MAIC. Based on these meetings, we identified that we required participants from MAIC as they represent the governing body for the QLD CTP claims process, and ND to understand the day-to-day execution of the insurance claim process. Based on

¹ Funding provided by MAIC (<https://maic.qld.gov.au/>).

this information, we performed purposeful sampling (Flick 2018) identifying two participants each from MAIC and ND. During interviews, we identified best practice may vary due to claim complexity. Therefore, we conducted an additional interview at ND to examine less complex claims (complex claims had already been captured adequately in prior interviews). We also identified the importance of lawyers in the process. As a result, we conducted two additional interviews with defendant lawyers. The importance of the claimant was also highlighted, and while we could not interview claimants due to confidentiality, we were able to gain insights into the claimant perspective through reviewing archival data (i.e., de-identified customer satisfaction surveys provided by the case organization). In total, we interviewed seven participants, two were from the legislator (MAIC), three from the insurer (ND), and two defendant lawyers.

The interviews were analyzed using the grounded theory techniques detailed in Sect. 3. In performing the analysis, through constant comparison of the open codes, we identified 12 themes (shown in Table 1) related to best practice. Through theoretical coding, which involved understanding the relationships between the themes, we identified that higher level themes - termed performance dimensions – were manifested by lower level themes, termed variables.

5.1.2 Challenges

C1: Covering Relevant Perspectives Identifying dimensions of best performance for insurance claims processing is challenging because different perspectives of the process are held by the stakeholders (e.g., claimants, insurers, law firms, governing bodies) who have different, and sometimes opposing, objectives. For example, the insurer usually wants to keep the payout low, whereas the claimant usually wants to maximize payout. Best practice, overall, should provide a fair pathway to balance the objectives of involved stakeholders. Thus, we needed to ensure that all (or at least most) perspectives on best practice in the domain were covered. This was achieved through purposeful (Newman et al. 2013) and theoretical sampling (Glaser 1998).

C2: Decomposition of Performance Dimensions Decomposing performance dimensions into variables is challenging, as through performing constant comparison with literature, which is a central tenet of grounded theory, some performance dimensions identified had not been previously investigated in literature. In addition, the performance dimensions apparent in literature had been manifested in different ways by different variables when compared to the performance dimensions evident in this

Table 1 Defining the dimensions of best practice CTP claims management

Item	Dimension	Definition
PF	Process fairness	The extent to which the outcome of a legitimate claim is perceived as fair in terms of both compensation and access to rehabilitation
PF1 ^{L+}	Compensation Fairness	The extent to which a claimant's settlement is appropriate based on injury severity and economic loss
PF2 ^{L+}	Rehabilitation Appropriateness	The extent to which a claimant receives rehabilitation in a timely manner
PQ	Process quality	The extent to which the claims management process is handled in an equitable manner with transparent communication
PQ1 ^{L+}	Equitable treatment	The extent to which all claimants are treated equally regardless of demographic differences or legal representation
PQ2 ^{L-}	Communication transparency	The extent to which the insurer clearly explains the claims process and outcomes to the claimant
PQ3 ^{L-}	Liability Determination	The extent to which the necessary evidence is collected to reasonably determine whether the insurer is responsible for handling the claim
PQ4 ^{L-}	Investigation Appropriateness	The extent to which the necessary evidence has been collected to justify the settlement
PC ^{O+}	Process costs	The sum of the costs associated with claims handling
PT	Process timeliness	The extent to which the stages within the claims management process meet legislative requirements
PT1 ^{O+}	Process timeliness – Legislative	The extent to which the stages within the claims management process meet legislative requirements
PT2 ^{O+}	Process timeliness – Overall	The overall time taken from notification to finalization of claim

O: variable is observable, L: variable is latent, +: variable has been operationalized, -: variable has not been operationalized

case. Theoretical coding, ensuring relevant perspectives were covered, and performing member-checking helped overcome existing limitations in literature to develop a robust understanding of best practice dimensions and their variables.

5.1.3 Results

The analysis of the interview data (using grounded theory techniques) identified four dimensions and their constituent variables which were either latent or directly observable. The **four dimensions** of best performing claims are: process fairness, process quality, process costs, and process timeliness (see Table 1).

While process fairness proved the most salient dimension and focused on the outcomes of the process, many respondents also emphasized process quality. Process quality takes into account equitable treatment of claimants and the need to transparently communicate with claimants. Process quality consists of four variables: equitable treatment, communication transparency, investigation appropriateness, and liability determination appropriateness. Process timeliness and process costs were also considered core dimensions. Process timeliness consists of two variables to capture whether deadlines imposed by legislation are met, and whether the overall claim (from start to finish) is handled in a timely manner. Process costs includes all costs associated with managing the claim (e.g., legal, investigations related to liability determination and investigation appropriateness, and rehabilitation management) excluding settlement and rehabilitation services. No specific thresholds were discussed for process costs. However, participants acknowledged the need to minimize costs to ensure CTP remains feasible.

5.1.4 Lessons Learned

Our findings highlight that a solely deductive approach to analyzing best practices is insufficient as key dimensions will be overlooked. Thus, having a good understanding of the domain is paramount and the dimensions should be inductively identified. In addition, the dimensions are complex and can be manifested by multiple variables.

Different participants in the process, may ultimately have different perspectives as to how these variables are defined and evaluated. As such, it is important to interview multiple process participants across diverse roles to ensure a complete view of the process is captured to identify the appropriate dimensions of best practice in a given domain. Semi-structured interview protocols should be used to guide the interviews, but the interviewer needs to follow up with themes (e.g., dimensions, conditions) as they emerge. As such, the initial pool of participants, as identified from

purposeful sampling, should not restrict the participant pool. Rather additional interviews will need to be conducted to ensure no relevant perspectives are overlooked.

5.2 Stage 2: Recognizing and Isolating Best Performing Cases

5.2.1 Objective and Methods Used

The main objective of Stage 2 is to identify cases that can be considered as best performing cases from historical data in order to learn about best practice process behavior.

We conducted a detailed analysis of historical data provided by MAIC. We analyzed 31570 claims finalized between 2013 and 2018 by five insurers. The CTP Scheme Milestone dataset covers nine major milestones in the processing of claims. For each claim the date each milestone was reached is stored, together with up to 58 attributes describing the claim context, such as demographic information of the claimant, nature and severity of injury and data relating to legal representation. These attributes are then analyzed to identify appropriate variables to measure each performance dimension.

Stage 2 consisted of four steps. First, we computed the values of the performance measures presented in Table 1. Using the dataset provided by MAIC we quantified six of the identified variables. The first Process Timeliness measure (PT1) relates to legislative guidelines as described previously. The second Process Timeliness measure (PT2) is calculated as the difference of days between the notification and finalized dates to measure overall processing duration. Processing cost (PC) summarizes the processing costs for a claim. The Compensation Fairness measure (PF1) is calculated as the difference between general damages scheduled for an injury scale value (ISV) according to the Civil Liability Regulation 2014 (Queensland Government 2014), and the actual amount paid. The Rehabilitation Fairness measure (PF2) is computed as number of days until access to rehabilitation is provided to a claimant. The most complex measure is that of Equitable Treatment (PQ1) which relates to how claimants are treated across different claims. Firstly, we compute the median across (normalized values of) PT1, PT2, PC, PF1 and PF2 for all cases. The value of PQ1 for a particular case is then the difference between the vector of measurements for this case and the vector of medians. A data-related challenge emerged as not every measure can be operationalized using the data provided. Secondly, we split cases that are expected to behave in a similar way into clusters and then classify the performance of cases within the same cluster. We worked closely with the stakeholders to identify meaningful cohorts of similar cases (e.g. legally

represented vs. direct claimants, different injury severity levels) for use in **clustering**.

Thirdly, we computed individual measures and then computed the normalized values of all measures for each dimension. An overall ranking of claims by “goodness” of practice was achieved by using a weighted average of the values of all dimensions, using equal weights for all dimensions. However, depending on the domain, stakeholders can use the weights to prioritize certain dimensions.

Using this aggregated measure, cases were binned into different categories and labeled as such (i.e., the top 25% of cases as high-performing cases; the middle 50% as medium performing and the bottom 25% as low-performing cases). A second challenge related to *binning* was encountered due to the nature of characteristics of the process analyzed. We also clustered performance dimensions using different clustering methods (Brock et al. 2011) and different numbers of clusters and evaluated the resulting clustering structures using the Silhouette index (Starczewski and Krzyżak 2015).

5.2.2 Challenges

C3: Operationalizing Latent Variables While certain variables related to the time and cost dimensions are usually straight forward to measure (i.e., observable), others relating to fairness and quality are less so. We carefully examined all available data values in the data set and determined values that could be used to measure three of the six sub-dimensions after discussions with stakeholders. We found that there is little data present in the given data set to compute three variables on Process Quality (i.e., communication transparency, liability determination, investigation appropriateness). Thus, these variables could not be operationalized with the given data set.

C4: Meaningful Binning This challenge was encountered due to the distribution of claims on the performance spectrum. The vast majority of claims were near the positive ends of the performance dimensions with a small but long tail of badly performing cases. We tested multiple approaches to determine the best way forward. Firstly, a threshold approach was used where the highest ranked 25% of cases were classified as high performing, the next 50% of case as moderately performing and the remaining 25% as low performing. A Data Envelopment Analysis (DEA) approach (Bogetoft and Otto 2010) was used but the result was very similar to the outcome of the binning approach described above: 20% of claims were labeled as efficient and efficiency scores of many moderately performing claims were high.

5.2.3 Results

We computed the individual measures and normalized values of all measures for each dimension. An overall ranking of claims by “goodness” of practice was achieved by using a weighted average of the values of all dimensions, using equal weights for all dimensions. Using this value, cases were binned into different categories and labeled as such (i.e., the top 25% of cases as high-performing cases; the middle 50% as medium performing and the bottom 25% as low-performing cases). These examples (see Table 2) also show, that the difference in performance between high performing cases and medium performing cases is very small although differences in one dimension can be significant. It can also be seen that low performing cases perform badly on multiple dimensions. Figure 2 shows the fractional distribution of low, medium, and high performing cases across the five CTP insurers (ND plus 4 commercial insurers).

Clustering results are consistent with the outcomes of other binning methods we applied: they show that there are significant performance differences between low-performing claims (approximately 10%) and other claims (hence, they are separated into two clusters); however, other claims (i.e., medium-performing and high-performing) are included in one cluster (which indicates that performance differences for these claims are not significant). This meant there was little difference between our top and average performing cases, which would make it more difficult to pick up strong differences between the performance classes.

5.2.4 Lessons Learned

It is important to have a good understanding of both the domain and the data to be able to operationalize measures for latent variables. It is critical to involve stakeholders closely in the development of proxy measures for these variables. Furthermore, the distributions of attributes in the data will severely affect the results of mining techniques used in the approach. It is therefore critical to determine with stakeholders the key attributes for clustering and to use appropriate binning approach early on to minimize confounding factors in the analysis.

5.3 Stage 3: Learning Best Practice From Best Performing Cases

5.3.1 Objective and Methods Used

The objective of Stage 3 is to learn best practice process behavior from data.

Table 2 Computed measures, dimensions and performance classifications - example cases

Case	Measures											Performance Category
	Process timeliness 1 (Days)	Process Timeliness 2 (Days)	Process Cost (\$AUD)	Compensation Fairness (\$AUD)	Rehab Appropriateness (Days)	Equitable treatment	Process Timeliness	Process Cost	Process Fairness	Process Quality	Performance	
1	0	67	213	0	1	.0636	1	.9992	.9979	.9670	.9924	High
2	0	239	326	0	1	.0632	.9867	.9988	.9979	.9673	.9897	High
3	48	712	8585	0	3	.1445	.8803	.9697	.9936	.9022	.9479	Med
4	0	3004	125	0	0	.0646	.7737	.9996	1	.9662	.9479	Med
5	0	857	117311	37480	13	1.0435	.9391	.5856	.4725	.1829	.5305	Low
6	223	4076	117206	600	210	1.1475	.3660	.5860	.5471	.0997	.4292	Low

We analyzed data from two sources. The first is the same CTP Scheme Milestones dataset used in Stage 2. The second dataset includes more detailed process data from the *Nominal Defendant* (ND), recorded for claims finalized between 2012 and 2018. The ND claims processing dataset contains more detailed (compared to CTP Scheme Milestone dataset) data of the claims handling process at the Nominal Defendant. It contains records of each activity performed by the ND, with an overall 51 unique activities that can occur during the processing of a claim. In the analysis, we used the open source process mining framework ProM², which provides functionality to load, filter and transform event logs and has many plugins to enable all kinds of process analysis.

When analyzing cases from all insurers for contextual differences between high-performing and low-performing claims we, firstly, used supervised machine learning techniques (Aggarwal 2015) on context data to identify correlations between claims context attributes and performance outcomes of claims. We used decision trees, decision forests, logistic regression and treatment learning, but none of the techniques produced strong models with the given attributes. In both datasets, a small number of claims had to be dropped from the analysis due to not aligning with the process model. We, furthermore, undertook a performance analysis (Wynn et al. 2017) comparing processing times for different parts of the process across all three performance categories. Finally, we applied process discovery (Leemans et al. 2014) and process analysis (Mannhardt et al. 2015) techniques to highlight differences in process behavior between high performing and low performing cases in the ND dataset.

5.3.2 Challenges

C5: Availability of process data behavior A challenge with the analysis of the CTP Scheme Milestone dataset was that while there is lot of contextual data available, the behavioral data is at a very high level. This means that much of the behavior in processing the claim is actually not in the data, thus the approach cannot identify detailed behavior from the data.

C6: “Curse of Dimensionality” As many types of ND claims have distinctive behaviors we separated them into clusters of similar behavior. However, this resulted in many clusters too small for analysis, meaning a lot of data could not be used in the analysis. It is therefore important to identify critical differences between claims but also to balance case similarity with reasonable cluster sizes. We regularly interacted with stakeholders to identify attributes relevant for clustering.

² <http://www.promtools.org>.

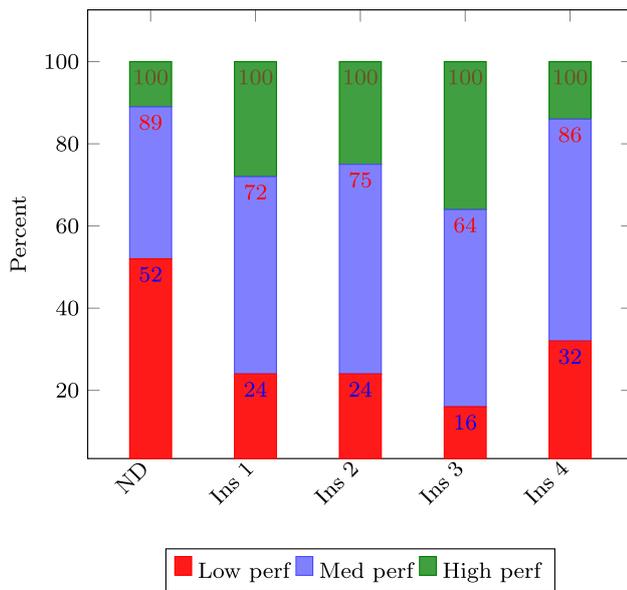


Fig. 2 Distribution of low/medium/high performing cases among five insurers

C7: Unstructured Process ND claims challenged one of our underlying assumptions. We assumed process behavior would have a distinctive effect on the performance measures. While we found some evidence of this, contextual factors such as the type of injury have a large effect on claim performance in multiple dimensions. It can be difficult to quantify such impacts from the data alone, so consequently it is challenging to disentangle effects from the type of case and the processing behavior of the case.

5.3.3 Results

Results will be presented first for the CTP Scheme Milestones dataset and then for the ND Claims dataset.

CTP Scheme Milestones Analysis Figure 3 shows a high-level Petri net model of the process based on the milestone data. As can be seen, behavioral choices that occur in the process depend on whether a solicitor is engaged, whether there was a conference and whether the case was settled. Another less obvious behavioral choice is whether a solicitor has been engaged before notification or after notification but before conference. The analysis on context differences indicated that whether the case went to litigation or not is a moderately good indicator for a case performing poorly.

In order to analyze the impact of these process decisions we annotate the milestone dataset with two binning criteria (10/80/10 and 25/50/25).

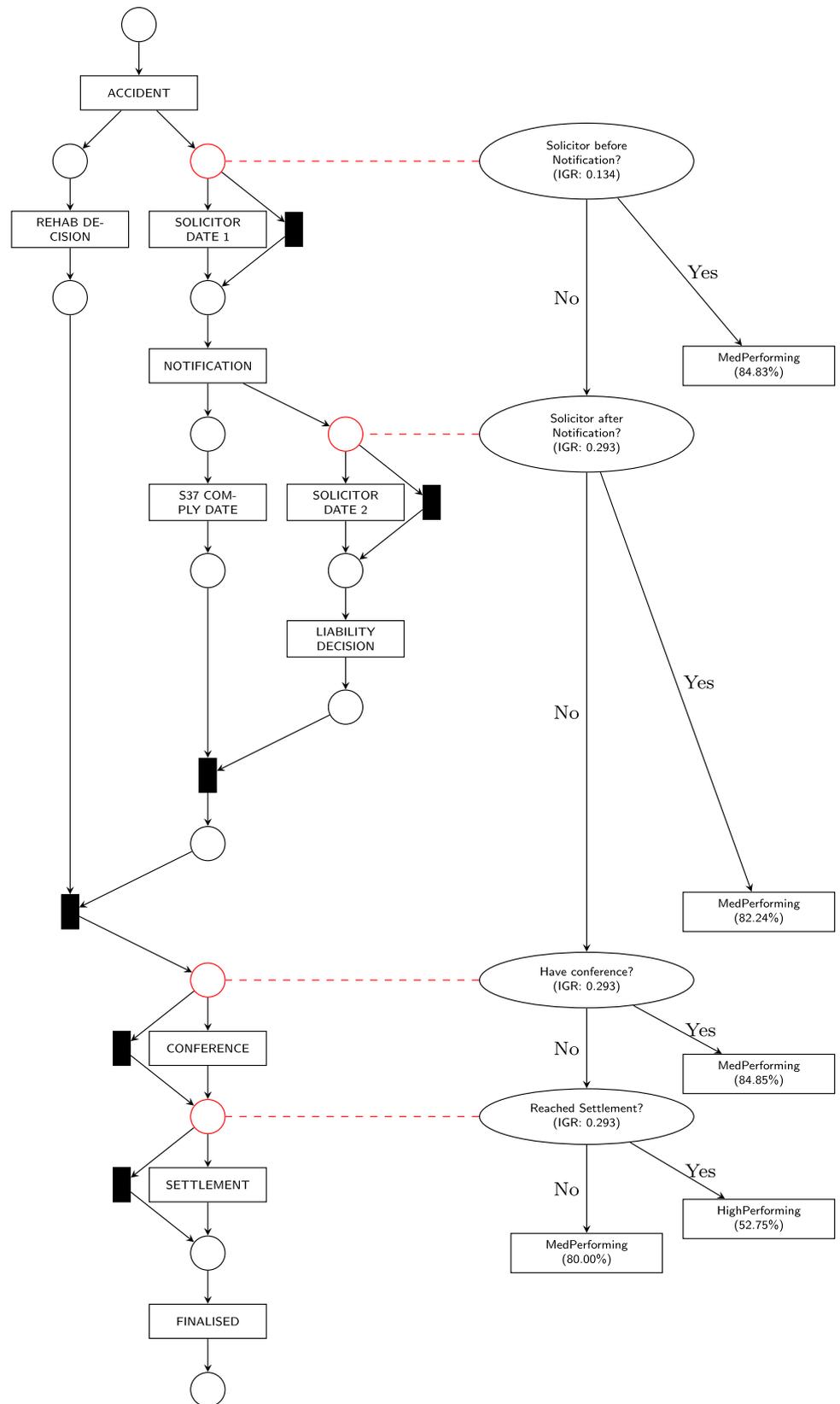
Using the 10/80/10 split, we created a vector for each case with attributes representing the behavioral choices in the case, i.e., whether a solicitor was engaged before

notification or after notification but before conference; whether a conference was chosen; and whether a negotiated settlement was arrived at. Each case was labeled according to its performance. Using this data set, we trained a decision tree with selection of attributes for splitting the tree based on IGR (Information Gain Ratio (Quinlan 1986)) values to see whether some of the behavioral factors could explain differences between claims in the different performance classes. The resulting decision tree (shown in Fig. 3 with case attributes lined up with decision points in the process model) had reasonable accuracy (80.77%) and showed that claims that were settled without solicitor and conference tend to be classified as high performing. The tree, however, does not explain low performing claims. Using a treatment/contrast set learning approach on both the high-performing class and the low-performing class individually provided additional detail. The treatment learner confirmed the indication of high performance for cases that get settled without solicitor and conference. Treatments for the low-performance class indicate that having a conference increases the likelihood of being classified as low performance. Similarly, having a solicitor increases the likelihood of ending up in the low-performing class.

Regarding differences in performance between high and low performing cases, the differences across cohorts are consistent across all activities. This means low performing claims always take longer than medium performing claims, which always take longer than high performing claims, although the amount of difference varies. Early on in the process the difference is small and it is most pronounced for the conference activity. The performance of cohorts is therefore consistent across all stages of the process, i.e., there is no single activity or stage where low performance claims are bottlenecked.

ND Claims Analysis We first split ND claims based on values of several attributes which are known to affect claim performance (age, injury severity, legal representation and vehicle category). We excluded from the analysis claims for claimants under the age of 18 as such claims follow a different process. Other claims were grouped based on values of the three attributes: legal representation (whether a claim was legally represented or not), vehicle category (whether a vehicle at fault was identified or not) and injury severity (four categories). The resulting claim clusters and the number of claims in each cluster are shown in Table 3 (clusters in the table are sorted by size). The majority of clusters only included few cases and were not further analyzed. Performance was evaluated for four largest clusters (with the number of claims ranging from 105 to 279). The selected finalized cases included 147 high-performing cases, 176 cases with medium performance and 42

Fig. 3 Process model for CTP milestone data (left) with process choices (highlighted red) and decision tree for process choices (right)



low-performing cases. The process data contained 51 activities and all claims followed unique process paths.

We applied decision tree classification (using activity executions in cases as descriptive variables) and identified activities that were associated with high-performing or low-performing cases. The event log was filtered to include four such activities (Fig. 4) and also activities “Upload new claim” and “Finalized” which are performed in all cases. We selected cases that follow five most frequent process paths for high-performing cases (80%) and for low-performing cases (86%) and discovered a process model from the resulting event log using the “Inductive miner” (Leemans et al. 2014) ProM plug-in, which can discover a fitting process model (i.e., a model “able to reproduce all observed behavior”) that is structured and sound (“free of deadlocks and other anomalies”) (Leemans et al. 2014). The selected high-performing and low-performing cases were replayed on the discovered process model using the “Multi-perspective Process Explorer” (Mannhardt et al. 2015) ProM plug-in; the results are depicted in Fig. 4.

Figure 4 shows the main process differences between high-performing and low-performing cases. In high-performing cases activity “Appoint Legal Panel” was not performed and conference was held in only few cases (6.8% as further investigation showed), while both activities were performed in most low-performing cases. Settlement was performed in all low-performing cases but was not performed in many high-performing cases (43.6%); while activity “S37(4) Notice” is more frequently performed in high-performing cases than in low-performing cases. Stakeholders commented that these findings are consistent with their expectations: the conference and the appointment of the legal panel in a claim indicate a higher complexity of the claim, which is expected to have a higher cost and a longer processing time and cases without settlement are expected to be faster and cheaper.

We conducted further analysis using individual dimensions. We binned the claims using the value of the individual dimension rather than the overall performance value. We then compared activity frequencies in high-performing and low-performing claims for each dimension, the results are depicted in Fig. 5. We can see that activities “Conference” and “Appoint Legal Panel” are more frequent in low-performing claims for all dimensions; however, the difference is more pronounced for the cost dimension. Although these activities typically happen late in the process and are often associated with more complex cases, the occurrence of these activities in a claim is not a very strong indicator of poor performance for the time dimension. Figure 5 also shows that while the execution of activity “S37(4) Notice” is more often associated with high-performing claims for the time, cost and fairness

dimension, the activity is performed more frequently in low-performing claims for the quality dimension. On the other hand, activity “Settlement” is more frequent in low-performing claims for the time, cost and fairness dimension; however, for the quality dimension it is more frequent in high-performing claims.

The analysis of individual dimensions helped uncover additional insights about process differences between high-performing and low-performing claims. We learned the differences can be significant for some performance dimensions (e.g., the execution of activities “Conference” and “Appoint Legal Panel” for the cost dimension) and negligible for others. We found some process behavior can be an indicator of high performance for one dimension and poor performance for another dimension (e.g., the execution of activity “S37(4) Notice” is associated with better performance with respect to time and cost, but not quality).

5.3.4 Lessons Learned

When a data-driven approach to identifying best practice is used it is important that behavioral data (i.e., process data) exists at the right level of detail. Data that only contains high-level milestones provides very limited details on the behavior that leads to the performance outcomes observed. For example, when a claim is seen to take a long time from liability decision to settlement, we can not identify the reason as much of the communications and negotiations happening between claimant, insurer and their legal representations that consume this time are not captured in this data.

In the presence of highly unstructured processes, process mining techniques can struggle to identify differences between well performing and badly performing cases. Using highly filtered event logs or high-level process features such as the presence of specific activity executions rather than full process models can help to identify differences in such scenarios. It is important to carefully validate these findings with stakeholders to check that the results are still representative of the process in general.

6 Discussion and Limitations of the Study

Table 4 summarizes key challenges encountered in various stages of the study discussed in Sect. 5.

In summary, we observe that a mixed-method approach to determining a best practice process can work well. The qualitative interviews can identify domain-specific dimensions of best practice that extend beyond the usual Devil’s Quadrangle measures. For our case study, it resulted in highlighting fairness as a major objective in injury compensation claim processing. The quantitative

Table 3 Clustering of ND claims

#	Legally represented	Identified vehicle	Injury severity	Cluster size
1	Yes	No	Category 1	279
2	Yes	No	Category 2	238
3	Yes	Yes	Category 1	218
4	Yes	Yes	Category 2	105
5	No	No	Category 1	39
6	No	Yes	Category 1	25
7	No	No	Category 2	24
8	Yes	No	Category 3	20
9	Yes	Yes	Category 3	12
10	No	No	Category 3	9
11	Yes	No	Category 4	5
12	No	Yes	Category 2	4
13	No	Yes	Category 3	3

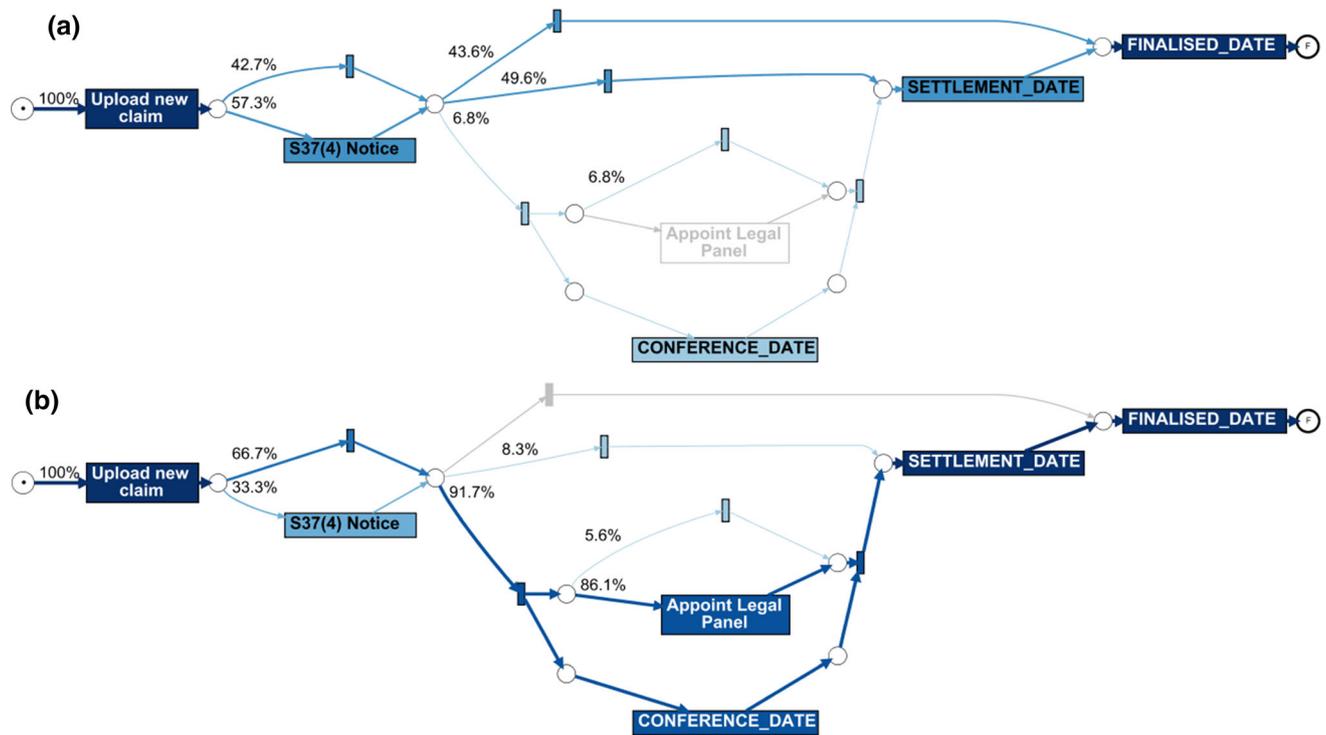


Fig. 4 Main process differences between high-performing **a** and low-performing cases **b**

analyses can explicate the gap between the measures we can easily capture from data and the measures of concepts that are highly relevant to stakeholders but hard to capture using existing data. Thus, the approach is found to be effective at integrating domain knowledge from stakeholders (e.g. in the chosen measures and the features to

define cohorts by) with evidence from the data. This means that the approach is more robust against subjectivity from stakeholders that qualitative approaches suffer from (Dumas et al. 2018) as well as being traceable in terms of how results are derived quantitatively so that stakeholders can

Fig. 5 Main process differences between high-performing (green) and low-performing cases (red) for individual dimensions: the percentage of claims in which a given activity was performed

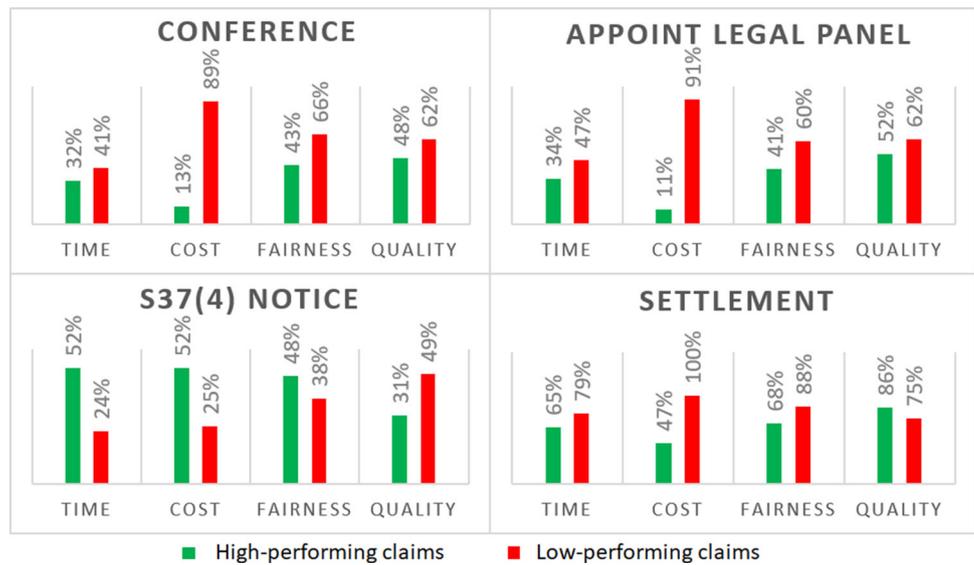


Table 4 Summary of challenges and reasons, methods employed to address challenges, and lessons learned

Challenge [Reason]	Method	Lessons learned
C1:Covering relevant perspectives required [Chosen approach]	Preliminary meeting with stakeholders to identify perspectives; purposeful sampling of interviewees from each perspective; review of archival data as a proxy for unavailable interviewees	It is important to interview multiple process participants across diverse roles to ensure a complete view of the process; purposeful sampling to select initial pool of interviews; expand interview pool as necessary to cover new perspectives
C2:Decomposing Dimensions [Chosen approach]	Inductive analysis using grounded theory techniques was used to identify the relevant performance dimensions	Through constant comparison with literature, it became evident that best practice performance dimensions extend beyond literature and a deductive approach to identifying best practices is, therefore, insufficient
C3:Operationalizing latent variables [Process/data characteristics]	Thorough examination of complete dataset for attributes that contribute to variables; stakeholders involved in determining relevant data attributes from which proxy measures were derived	Variables identified by stakeholders from interviews will rarely line up with data from IT systems; involve stakeholders in developing proxies from available data together with clustering to ensure dissimilar cases are not considered in the same cohort
C4:Meaningful binning [Case study data]	Multiple binning strategies (including clustering); normalization and weighting of dimensions to achieve case separation	Involve stakeholders in determining key attributes for clustering; Depending on the domain, stakeholders can use the weights to prioritize certain dimensions
C5:Availability of process data [Case study data]	We constructed process models using a combination of automated discovery techniques and manual adjustments based on domain knowledge	When a data-driven approach to identifying best practice is used it is important that behavioral data (i.e., process data) exists at the right level of detail
C6:Curse of dimensionality [Process/data characteristics]	Analyze distribution of individual attributes and bin based on these values rather than overall case performance; Compare activity frequencies in high-performing and low-performing claims for each dimension	It is important to identify critical differences between claims but also to balance case similarity with reasonable cluster sizes. Interact regularly with stakeholders to identify attributes relevant for clustering
C7:Structuredness of process [Process/data characteristics]	Where process is unstructured or shows high number of variations, using highly filtered event logs or high-level process features such as the presence of specific activity executions rather than full process models can help to identify differences in such scenarios	Contextual factors can be the root cause of variations; it is difficult to disentangle effects due to the type of case from those due to the processing behavior of the case; it is important to carefully validate these findings with stakeholders to check that the results are still representative of the process in general

validate if results are sensible to avoid results that come from erroneous interpretations of the data.

As with most data analysis techniques, the success of this approach is highly dependent on the available data. Our approach relies on mining differences in behavior from cases that have similar context and prescribed behavior but differ in terms of achieving performance outcomes, which means cases have to be clustered into cohorts to be analyzed separately. If there are many dimensions in context and prescribed behavior, then even large data sets are reduced to small clusters of cases that provide little power for analysis.

The applicability of the proposed approach to identify best practice in other domains depends on several assumptions. Centrally, we assume that there exists ‘best practice’ *process behavior* for the domain under investigation and that the effect of such behavior on one or more performance measures is larger than that of random noise. Furthermore, the use of process mining techniques to identify behavioral differences relies on the fact that ‘best practice’ is in the form of process behavior and is, consequently, logged in the form of activities in the data. If these assumptions do not hold for a particular process under examination, the approach may not be suitable.

In terms of improvement, we believe that integrating an additional qualitative step to Stage 3 of our approach to uncover certain best practice behavior that has not been **explicitly** recorded in the data, or for which supporting data is missing, can broaden and deepen findings from data analysis. Such an additional step would be designed to cater for data that may be helpful in identifying best performing cases, but which does not directly contribute to the identified performance measures. In addition, to better identify performance measures, best performing cases, and overcome limitations of missing event data, the event log can be triangulated with additional data sources.

7 Conclusions

In this paper, we presented a detailed case study that applied a mixed-method approach to identify ‘best practice’ in motor accident injury compensation claims processing. We first identified four dimensions of best practice in the domain of CTP claims processing and nine sub-dimensions. We then analyzed historical claims data captured from all QLD CTP insurers and were able to quantify measures for six of the nine sub-dimensions. We then analyzed claim performance across different insurers and showed differences in the distribution of high and low performing claims both by insurer and over time.

Next we identified key differences in context, behavior and performance between high and low performing claims

to extract ‘best practice’. In the context factors we picked up some correlation between claim costs and performance and whether the claim went to litigation. In process behavior reaching settlement without using a solicitor or conference was a good predictor for high performing claims while having a conference was a good predictor of low performing claims. In the performance analysis we found that no single bottleneck seems to exist as lower performing claims perform worse across all milestones rather than one particular milestone. In order to evaluate our approach with more detailed data we analyzed the Nominal Defendant data in a similar manner. Despite the low number of claims we were able to identify process behavior differences between high and low performing claims.

Overall, we demonstrated that the proposed mixed-method approach to mine best practice from data works in the real-world context, but has fairly steep requirements on the quality and detail of data that needs to be available to perform the analysis.

Acknowledgements The work presented in this paper was funded by a grant from the Queensland Motor Accident Insurance Commission (MAIC). We gratefully acknowledge the contributions of Dr Suriadi Suriadi (Business Process Management group, Queensland University of Technology) to the project on which this paper is based.

References

- Aggarwal CC (2015) Data mining: the textbook. Springer, Heidelberg
- Andrews R, Wynn MT, ter Hofstede AHM, Xu J, Horton K, Taylor P, Plunkett-Cole S (2018) Exposing impediments to insurance claims processing. Business process management cases. Springer, Heidelberg, pp 275–290
- Bogetoft P, Otto L (2010) Benchmarking with DEA, SFA, and R, vol 157. Springer, Heidelberg
- Bose J, Mans R, van der Aalst W (2013) Wanna improve process mining results? It’s high time we consider data quality issues seriously. In: CIDM, IEEE, pp 127–134
- Brand N, Van der Kolk H (1995) Workflow analysis and design. Kluwer Bedrijfswetenschappen, Deventer
- Brock G, Pihur V, Datta S, Datta S, et al. (2011) clValid, an R package for cluster validation. J Stat Softw
- Cho M, Song M, Comuzzi M, Yoo S (2017) Evaluating the effect of best practices for business process redesign: an evidence-based approach based on process mining techniques. Decis Support Syst 104:92–103
- Christmann P (2000) Effects of “best practices” of environmental management on cost advantage: The role of complementary assets. Acad Manag J 43(4):663–680
- Creswell JW, Miller DL (2000) Determining validity in qualitative inquiry. Theor Pract 39(3):124–130
- De Leoni M, van der Aalst WM (2013) Data-aware process mining: Discovering decisions in processes using alignments. In: Proceedings of the 28th annual ACM symposium on applied computing, pp 1454–1461
- De Leoni M, van der Aalst WM, Dees M (2016) A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. Inf Syst 56:235–257

- De Smedt J, Hasić F, vanden Broucke SK, Vanthienen J (2017) Towards a holistic discovery of decisions in process-aware information systems. International conference on business process management. Springer, Heidelberg, pp 183–199
- De Weerd J, Schupp A, Vanderloock A, Baesens B (2013) Process mining for the multi-faceted analysis of business processes - a case study in a financial services organization. *Comput Indust* 64(1):57–67
- del Rio-Ortega A, Resinas M, Cabanillas C, Ruiz-Cortes A (2012) On the definition and design-time analysis of process performance indicators. *Inf Syst* 38(4):470–490
- Dumas M, La Rosa M, Mendling J, Reijers HA (2018) Three process discovery challenges. In: *Fundamentals of business process management*, 2nd edn, Springer, Heidelberg, chap 5.1.2, pp 162–165
- Flick U (2018) *An introduction to qualitative research*. Sage, Thousand Oaks
- Francis C, Iglesias M, Walsh J (2009) Accident compensation claims management-lessons learnt and claimant outcomes. Presented to the Institute of Actuaries of Australia 12th Accident Compensation Seminar 22–24 November 2009, Melbourne. https://www.actuaries.asn.au/Library/ACS09_Paper_Francis%20et%20al._Accident%20Compensation%20Claims%20Management.pdf
- Glaser B (1998) *Doing grounded theory: Issues and discussions*. Sociology Press
- Greenland S, Robins JM, Pearl J (1999) Confounding and collapsibility in causal inference. *Statist Sci* 14(1):29–46
- Kis I, Bachhofner S, Di Ciccio C, Mendling J (2017) Towards a data-driven framework for measuring process performance. In: Reinhartz-Berger I et al (eds) *Enterprise, business-process and information systems modeling*. Springer, Heidelberg. https://doi.org/10.1007/978-3-319-59466-8_1
- Leemans SJJ, Fahland D, van der Aalst WMP (2014) Discovering block-structured process models from event logs containing infrequent behaviour. In: Lohmann N, Song M, Wohed P (eds) *BPM workshops*. Springer, Heidelberg, pp 66–78
- MacKenzie SB, Podsakoff PM, Podsakoff NP (2011) Construct measurement and validation procedures in mis and behavioral research: Integrating new and existing techniques. *MIS Q* 35(2):293–334
- Mannhardt F, De Leoni M, Reijers HA (2015) The multi-perspective process explorer. *BPM (Demos)* 1418:130–134
- Mannhardt F, De Leoni M, Reijers HA, Van Der Aalst WM (2016) Decision mining revisited-discovering overlapping rules. International conference on advanced information systems engineering. Springer, Heidelberg, pp 377–392
- Mansar S, Reijers HA (2007) Best practices in business process redesign: Use and impact. *Bus Process Manag J* 13(2):193–213
- Myers MD, Newman M (2007) The qualitative interview in is research: Examining the craft. *Inf Organ* 17(1):2–26
- Newman I, Lim J, Pineda F (2013) Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. *J Mixed Methods Res* 7(3):243–260
- Queensland Government (2014) *Civil Liability Regulation 2014*. Queensland Government
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106
- Reijers H, Mansar S (2005) Best practices in process redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omega* 33(4):283–306
- Rondini A, Pezzotta G, Cavalieri S, Ouertani MZ, Pirola F (2018) Standardizing delivery processes to support service transformation: A case of a multinational manufacturing firm. *Comput Indust* 100:115–128
- Rozinat A, van der Aalst WM (2006) Decision mining in ProM. International conference on business process management. Springer, Heidelberg, pp 420–425
- Starczewski A, Krzyżak A (2015) Performance evaluation of the silhouette index. International conference on artificial intelligence and soft computing. Springer, Heidelberg, pp 49–58
- van der Aalst WM, Dustdar S (2012) Process mining put into context. *IEEE Internet Comput* 16(1):82–86
- van der Aalst WM, Van Dongen BF (2013) Discovering petri nets from event logs. In: Jensen K et al (eds) *Transactions on petri nets and other models of concurrency VII*. Springer, Heidelberg, pp 372–422. https://doi.org/10.1007/978-3-642-38143-0_10
- van der Aalst WMP (2016) *Process mining: Data science in action*. Springer, Heidelberg
- van Eck ML, Lu X, Leemans SJJ, van der Aalst WMP (2015) PM²: A process mining project methodology. *LNCS* 9097:297–313
- Venkatesh V, Brown SA, Sullivan YW (2016) Guidelines for conducting mixed-methods research: An extension and illustration. *J Assoc Inf Syst* 17(7):435–494
- Wynn MT, Poppe E, Xu J, ter Hofstede AHM, Brown R, Pini A, van der Aalst WMP (2017) *ProcessProfiler3D: A visualisation framework for log-based process performance comparison*. *Decis Support Syst* 100:93–108
- Wynn MT, Suriadi S, Eden R, Poppe E, Pika A, Andrews R, ter Hofstede AHM, (2019) Grounding process data analytics in domain knowledge: A mixed-method approach to identifying best practice. In: *Proceedings of the business process management forum*, Vienna. Springer, Heidelberg, pp 163–179