Pre-hospital Retrieval and Transport of Road Trauma Patients in Queensland
A Process Mining Analysis

Robert Andrews[1]([✉]), Moe T. Wynn[1], Kirsten Vallmuur[1],
Arthur H. M. ter Hofstede[1], Emma Bosley[2], Mark Elcock[3],
and Stephen Rashford[2]

[1] Queensland University of Technology (QUT), GPO Box 2434, Brisbane, Australia
r.andrews@qut.edu.au
[2] Queensland Ambulance Service (QAS), GPO Box 1425, Brisbane, Australia
[3] Retrieval Services Queensland (RSQ), GPO Box 1425, Brisbane, Australia

Abstract. Existing process mining methodologies, while noting the importance of data quality, do not provide details on how to assess the quality of event data and how the identification of data quality issues can be exploited in the planning, data extraction and log building phases of any process mining analysis. To this end we adapt CRISP-DM [15] to supplement the Planning phase of the PM^2 [6] process mining methodology to specifically include data understanding and quality assessment. We illustrate our approach in a case study describing the detailed preparation for a process mining analysis of ground and aero-medical pre-hospital transport processes involving the Queensland Ambulance Service (QAS) and Retrieval Services Queensland (RSQ). We utilise QAS and RSQ sample data to show how the use of data models and some quality metrics can be used to (i) identify data quality issues, (ii) anticipate and explain certain observable features in process mining analyses, (iii) distinguish between systemic and occasional quality issues, and, (iv) reason about the mechanisms by which identified quality issues may have arisen in the event log. We contend that this knowledge can be used to guide the extraction, preprocessing stages of a process mining case study.

Keywords: Process mining · Data quality · Pre-hospital · GEMS · HEMS

1 Introduction

Pre-hospital care and transport can be supplied by road services, aero-medical services or a combination of these two services. Comparing the various transport modes and escort levels etc. may lead to a better understanding of factors contributing to patient outcomes. However, there is limited research internationally examining the retrieval processes for patients from roadside to definitive care, and there has been no research conducted in the Queensland context.

This project, being conducted in collaboration with Queensland Ambulance Services (QAS) as providers of ground-based transport services, and Retrieval Services Queensland (RSQ) as coordinators of aero-medical transport services, aims to provide insights into a key question of interest to QAS and RSQ, i.e. "are we getting the right level of care to the right patients"? Specific questions to be considered include:

– Are we making the correct decisions about the dispatch of assets?
– Is it possible to validate the existing "45 min" guideline for transport to the nearest facility (and the by-pass to major trauma centre guideline)?
– Can the existing aero-medical request for launch protocol be validated?

It is proposed to address these questions using process mining techniques to:

– Discover the range of different care and transport processes undertaken for road trauma patients from roadside to definitive care
– Conduct conformance (to guidelines) and comparative performance analyses
– Identify key factors influencing deviance from standard care and delivery processes as given in the guidelines

The impact of data quality on process mining analyses is well recognised [5,12] and existing process mining methodologies [6,10] refer to taking into consideration characteristics and quality of the input data in the early stages of a process mining study. However, there is little guidance on how to actually assess process-data quality. We adapt the CRISP-DM (Cross Industry Standard Process for Data Mining as outlined in [15]) and apply it in our case study to (Fig. 1):
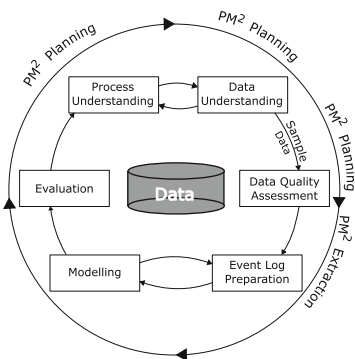
– Gain an understanding of the overall QAS and RSQ dispatch-retrieval-transport processes.
– Gain an understanding of QAS and RSQ data through development of data models and examination of sample data extracts.
– Conduct data quality assessments.
– Prepare event logs.
– Use the sample data to discover models of the individual QAS and RSQ retrieval/transport processes.
– Evaluate/conformance check the models.

Fig. 1. Our approach adapted from the CRISP-DM methodology outlined in [15].

The major contributions of this paper include:

– conceptual data models (Object-Role Models (ORM)) of data held by (i) a ground based ambulance service provider (QAS) and (ii) a coordinator of aero-medical retrieval and transport service provider (RSQ);
– an assessment of the quality (fitness for purpose) of the QAS and RSQ data for process mining analysis;

– a contribution to the knowledge on how to conduct a process mining study through a demonstration of the value of systematically identifying data-related issues prior to carrying out a process mining analysis; and
– a contribution to the knowledge-base in relation to Ground and Helicopter EMS dispatch processes in an Australian context.

2 Related Literature

We considered literature relating to process mining in the healthcare domain (and in pre-hospital transport in particular). We look also at process mining methodology and find that existing methodologies do not highlight the value of data quality assessment in the early stages of a process mining exercise. We found that little work has been done in applying process mining techniques to analyse pre-hospital processes.

2.1 Process Mining and Healthcare

Rebuge and Ferreira [10] propose a Business Process Analysis methodology for healthcare based on process mining. The methodology comprises: (1) the preparation of an event log; (2) log inspection; (3) control-flow analysis; (4) performance analysis; (5) organizational analysis; (6) transfer of results. While the methodology steps (1) and (2) are data-focused, consideration of the quality of data and its suitability for process mining is not considered. In [6] the authors propose PM^2 as a comprehensive, 6 step (Planning, Extraction, Data processing, Mining & Analysis, Evaluation, Process Improvement & Support) process mining methodology. In the description of PM^2, the authors do not mention event data-quality from the point of view of (i) informing the Extraction and Data Processing stages, (ii) possible impacts on Mining & Analysis. Mans et al. [9] discuss event data recorded in Hospital information Systems (HIS) and introduce the Healthcare Reference Model, a comprehensive data model designed to allow analysts to locate event data easily and to support data extraction. Rojas et al. [11] review 74 articles describing applications of process mining in the healthcare domain. Papers were characterised according to 11 points of relevance including process type, data type, frequently asked questions, analysis perspectives, tools, methodologies. The authors conclude that future work should focus on the implementation of process-aware hospital information systems along with improved visualisation and visual analytics techniques and an increased focus on conformance checking in case studies. Andrews et al. [2] discuss the application of process mining techniques in the analysis of healthcare process-related data focussing on data extraction, pre-processing and data quality assessment before considering challenges facing analysts in dealing with the semi-structured nature of healthcare processes when conducting discovery, conformance (comparative) performance analysis before providing some novel visualisation options. Little work has been done in applying process mining techniques to analyse pre-hospital processes. Lamine et al. [8] apply process mining and discrete event

simulation to assess the efficiency of emergency call centre operations in France and [3] apply process discovery, conformance checking and performance analysis in a case study involving ambulance services in Iran.

3   Case Study Description

3.1   High-Level Patients Retrieval/Transport Process in Queensland

Figure 2 is a high-level retrieval/transport model derived from operating guideline documents provided by QAS and RSQ. In Queensland, all emergency calls (calls to 000) are routed to a single, statewide call centre operated by QAS. The emergency centre operators gather as much information about the incident as they can from the caller reporting the incident. Usually, QAS will dispatch one or more ground-based ambulances to the scene of the incident, but may directly request aero-medical evacuation of injured person(s). Once on-scene, QAS paramedics (i) will provide first-level support to injured patients, (ii) may contact a senior on-call paramedic or QAS Medical Coordinator (an experienced emergency doctor) for treatment advice, and (iii) where the situation fits guidelines, may request aero-medical evacuation of injured person(s).

Where aero-medical retrieval/transport is required, the QAS Communications Centre Supervisor (CCS) calls the RSQ Communication Centre. The call is picked up by a QAS Emergency Medical Dispatcher (EMD) stationed at the dedicated Rotary Wing Desk within the RSQ Communication Centre. The EMD has access to the statewide QAS Computer Aided Dispatch (CAD) which shows
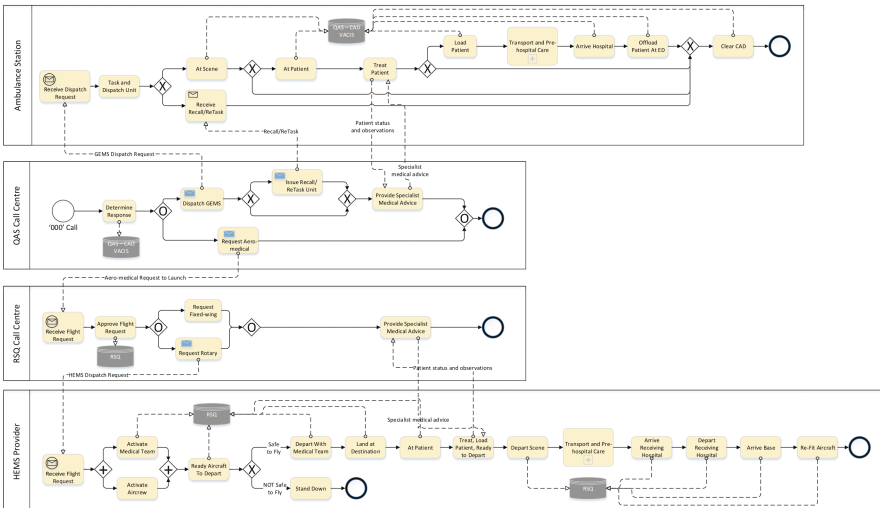

Fig. 2. BPMN model of emergency incident management - ground and aero-medical call centre, asset deployment and patient transport.

the Incident record. The EMD links the QAS CCS with the RSQ Medical Coordinator who discuss the incident and determine the optimal response. If the decision is made to dispatch an aircraft the EMD tasks the aircraft while the RSQ Medical Coordinator contacts the retrieval team to fly on the respective aircraft and provides the patient's clinical details. On arrival at the scene of the incident, or following contact with the patient, the retrieval team contacts the RSQ Medical Coordinator (via satellite phone or mobile phone). The RSQ Medical Coordinator provides specialist advice to, and oversight of, the retrieval team. They then determine the receiving hospital based on the patient's clinical needs and informs the receiving, on-duty Emergency Department Specialist of the incoming patient, their estimated time of arrival and their clinical condition and requirements. On arrival at the Receiving Hospital, the retrieval team hands-over to the Emergency Department Specialist.

3.2 Scenarios

From the process description, high level BPMN model, data models, and discussions with domain experts, it is possible to derive some scenarios which may play out in response to any incident:

1. Road-based response/s with treatment and no transport.
2. Road-based response/s with treatment and at least one primary transport.
3. Road-based response/s with treatment and a rotary wing primary transport.
4. Rotary wing inter-hospital (secondary) transfer.
5. Fixed wing primary transport.
6. Multileg primary transport (road + rotary or fixed wing).

In this preliminary study, only scenarios 1–3 are considered. The full study will consider all scenarios.

3.3 Data Models - Ground and Aero-medical Retrieval/Transport

From our understanding of emergency incident reporting-to-retrieval/transport (developed through interviews with domain experts, documentation describing QAS and RSQ data and informed by our literature review) we identified data relevant to the study that allows end-to-end traceability (notification to delivery to definitive care) and which allows segmentation of the data into cohorts of retrieval/transport cases of interest to the process stakeholders. The Object-Role Model [7] in Fig. 3 depict the main data attributes necessary to allow end-to-end traceability and case segmentation for QAS. A simlar model (not shown) was developed for RSQ. The main categories of data are as follows:

1. **Incident data** such as location of the incident, notification datetime the incident was reported to the emergency call centre and the priority of the incident.
2. **Patient data** including patient name, age, gender, pre-existing conditions, allergies, current medications and indigenous status.

3. **Transport data** which includes timestamped way-point data representing key case milestones, details of assessment of the scene, patient and injury by the paramedics, observations of the patient, management activities and procedures carried out by the ground-based paramedics or aircraft medical team, the destination hospital, and the patient outcome.
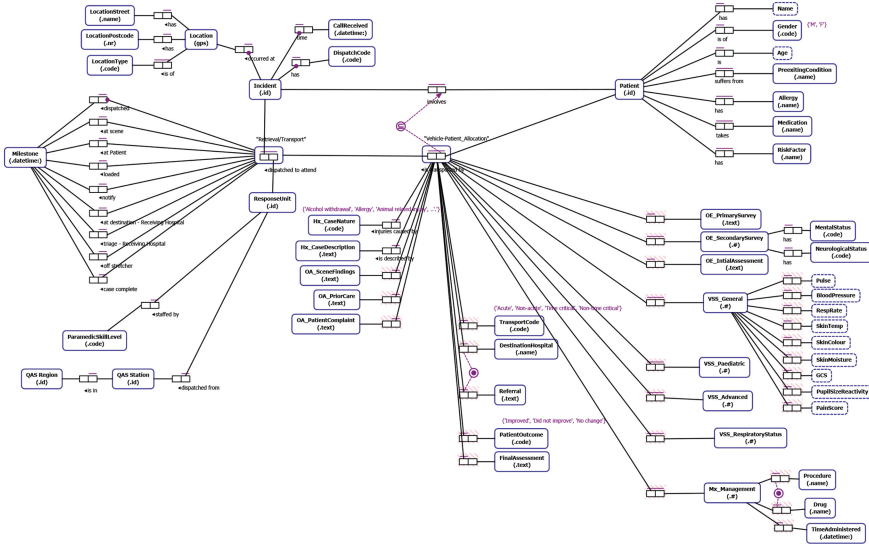

Fig. 3. ORM model of QAS data

4   Data Quality Assessment

Data quality is described as a multi-dimensional concept [13] with each dimension representing some (quantifiable) characteristic. For this study, we use 3 (Completeness, Precision, Uniqueness) of the 20 quality dimensions frequently mentioned in [14] and their associated metrics. The metrics were chosen as they provide insights into not only the state of the data in a particular column, but also into some possible impacts on process modeling. For instance, low values for the Precision metric [13] for datetime columns indicates coarse granularity (e.g. some values in the column may be at day level granularity). From a process mining perspective, this presents some issues in sequencing the events properly (day level granularity events will always appear to occur before milli-second level granularity events for events that have the same date). The Completeness metric [1] measures the fraction of the rows of the data set that have a value in the column. The Completeness metric then gives an indication of the suitability of the column for inclusion in an event log. For instance, if the column values are intended to be used to differentiate between cohorts of cases and the column is

only 25% complete, it will not be possible to properly segment the set of cases. Lastly, the Uniqueness metric [4] provides a measure of the similarity of values in the column. For datetime columns that represent event log times, it is often good to have high uniqueness, while for columns that represent activity labels, a certain degree of sameness is desirable. To conduct the assessment we (i) loaded the sample data into a relational database, then (ii) applied the column level quality checks, and (iii) checked for the presence of event log imperfection patterns as described in [12].

4.0.1 QAS Sample Data comprised a de-identified sample of 500 incidents attended by QAS between 01-July-2016 and 09-Jul-2016. The data set was compiled from two separate information systems maintained by QAS. The Computer Aided Dispatch (CAD) system records the datetime of incident notification, (first) vehicle assignment, vehicle arrival on scene, departure from the scene, arrival at destination (hospital) and finally completion of the assignment. Not all waypopint times are recorded for vehicles not involved in a patient transport. The Electronic Ambulance Report Form (eARF) records waypoint times for individual patients including vehicle en route, arrival at the scene, paramedics at the patient, patient loaded (for transport) and patient off-load (at hospital). Again, as not all attendances result in a transport, not all fields are populated.

The data was provided in tabular (Excel) format where each of the 15 columns represented an attribute of the attendance/transport (incident identifier, patient identifier and patient and vehicle waypoint times). The data set contained 12 datetime type columns with 2 waypoint times, one from the CAD system and one from the eARF, that likely represent the same event ('At Scene'). From the QAS process description, we note that there can be multiple units attending a single incident and multiple patients involved in a single incident. It is therfore possible to consider the data from at least three different "case" perspectives, i.e. an **incident** may be considered as a case, each **patient** may be considered as a case, or each **response unit** may be considered as a case. After consulting with the domain experts, it was determined that each patient should be the subject of the case. For the purposes of this part of the study, it was decided that eARF could be treated as surrogate patients, i.e. the eARF number would be the case identifier. Some of the time stamps are standardised across all records relating to a given incident to reflect the 'First Assigned' time (that is, all vehicles attending an incident will have the same value for the FIRST_ASSIGNED_CAD waypoint time). Others, such as On Scene/Depart Scene/Destination/Clear reflect the times for that specific unit. Not all timestamps are relevant to all attending units, hence some are empty e.g. D_LOADED_VACIS time isn't recorded for units not transporting a patient. After considerable cleaning and de-duping of the data it was possible to match 723 eARF (VACIS) records with response unit (CAD) records.

Table 1 provides values for three column-level metrics useful in assessing the quality of the de-duplicated data. Here we note that:

– the **Completeness** metric shows that only 3 of the date time columns are 100% complete which indicates that in any incident, not all patient and vehicle waypoints are completed. In particular, the 50% complete value for OFF_STRETCHER_VACIS indicates that only half the patients involved in incidents required road transport to hospital.
– the **Precision** metric (for datetime) values gives an indication of mixed granularity among the various timestamps.
– the **Uniqueness** metric gives an indication of the degree of distinct values found in the column. The FIRST_ASSIGNED_CAD value shows low Uniqueness indicating many repeated values. This reflects the QAS policy of assigning to all vehicles involved in an incident, the timestamp of the first vehicle assigned to attend the incident.

Table 1. QAS - Column-level data quality summary

Column name	Data type	Completeness	Precision	Uniqueness
T_INCIDENT	int	100%		100%
D_RECEIVED_CAD	datetime	100%	83%	100%
FIRST_ASSIGNED_CAD	datetime	100%	83%	58%
CAD (Vehicle) waypoints				
ON_SCENE_CAD	datetime	97%	83%	83%
DEPART_SCENE_CAD	datetime	57%	83%	82%
AT_DEST_CAD	datetime	57%	83%	82%
CLEAR_CAD	datetime	100%	83%	83%
eARF (Patient) waypoints				
EN_ROUTE_VACIS	datetime	94%	66%	84%
AT_SCENE_VACIS	datetime	95%	66%	86%
AT_PAT_VACIS	datetime	90%	66%	90%
LOADED_VACIS	datetime	52%	66%	89%
NOTIFY_VACIS	datetime	7%	66%	53%
OFF_STRETCHER_VACIS	datetime	50%	66%	93%

The distinctly different values of the Precision metric between the _CAD timestamps and _VACIS timestamps suggests a difference in granularity between the sets of timestamps. Investigation revealed that all the _VACIS timestamps were recorded at minute-level granularity while the _CAD timestamps were recorded at second-level granularity. The immediate effect of the mixed granularity on event ordering can be seen when considering two events that must, in reality, occur in a particular order, but which appear to happen in a different order (according to their timestamps). For instance, D_RECEIVED_CAD is the

time when QAS Call Centre is notified of an incident and D_EN_ROUTE_VACIS is the time a response unit is recorded as travelling to the incident scene. There are 52 (out of 723) cases where the D_EN_ROUTE_VACIS time is earlier than the D_RECEIVED_CAD time, and in 49 of these cases, the two timestamps are the same to minute-level granularity (as one example, for N_EARF = 76507098, D_RECEIVED_CAD = 2016-07-05 07:34:08 and D_EN_ROUTE_VACIS = 2016-07-05 07:34:00). This fits the description of the 'Inadvertent Time Travel' log imperfection pattern described in [12] and as such, if left unaddressed, has likely impact on process mining in terms of temporal ordering of events no longer matching reality, and incorrect activity/case durations and will likely result in discovered process models showing these two events in parallel rather than, as expected, in sequence. We note that there are also 15 cases where the value of D_CLEAR_CAD is earlier than the D_OFF_STRETCHER_VACIS time, however, this discrepancy appears to be due to some other mechanism (the difference between the two times is up to 1 h).

4.0.2 RSQ Data RSQ provided a de-identified sample of 500 aero-medical transports with case dates between 01-Mar-2017 and 28-Apr-2017 comprising 419 Inter-hospital Transfers, 78 Primary Response missions and 3 Search and Rescue missions. The data set was provided in tabular (Excel) format where each row represented a separate mission and each of the 128 columns represented an attribute of the mission. The data included 62 mission records where the Mechanism of Injury value was 'Vehicle accident' (comprising 35 Inter-hospital Transfers and 27 Primary Response missions). The data set contained only 12 datetime type columns. From a process mining perspective, this gives, at most, 12 different activities that can be extracted from the data. Table 2 provides values for some column-level metrics useful in assessing the quality of the data.

Table 2. RSQ - Column-level data quality summary for a sample of columns

Column name	Data type	Completeness	Precision	Uniqueness
SOURCE_ID	int	100%		100%
DATE_RETRIEVAL_REQUESTED	datetime	100%	33%	7%
TEAM_ACTIVATED	datetime	100%	65%	95%
READY_TO_DEPART	datetime	100%	65%	96%
DEPART_WITH_MEDICAL_TEAM	datetime	100%	65%	95%
LAND_AT_DESTINATION	datetime	100%	65%	96%
AT_SCENE_PATIENT	datetime	100%	65%	97%
DEPARTURE_READY	datetime	100%	65%	96%
ACTUAL_TIME_DEPART	datetime	100%	65%	96%
ARRIVE_AT_RECIEVING_HOSPITAL	datetime	100%	65%	96%
DEPART_RECIEVING_HOSPITAL	datetime	100%	65%	96%
ARRIVE_BACK_AT_BASE	datetime	100%	66%	93%
AVAILABLE_FOR_NEXT_TASKING	datetime	100%	63%	92%
MECHANISM_OF_INJURY	string	25%	100%	27%

Here we note that:

– the **Completeness** metric shows that all values are populated for the date
  time columns, while only 25% of the records in the log have a value for the
  MECHANISM_OF_INJURY column;
– the **Precision** metric (for datetime) values gives an indication of coarse
  granularity among the various timestamps. For instance, all values of the
  DATE_RETRIEVAL_REQUESTED column are day-level granularity, while
  all other datetime columns are at minute-level granularity.
– the **Uniqueness** metric gives an indication of the degree of distinct values
  found in the column. The DATE_RETRIEVAL_REQUESTED value shows
  low Uniqueness indicating many repeated values. This is not surprising
  given the narrow range of case dates (many cases on any given day). The
  SOURCE_ID column shows perfect uniqueness (every value different from all
  others), while Uniqueness value of 27% for the MECHANISM_OF_INJURY
  column is reflective of the value being populated from a limited set of allowed
  values (e.g. a pull-down on a form).

The datetime columns represent milestone events in a mission and are
expected to be sequential. We note that there are several violations of such
ordering apparent in the sample data. For instance:

Table 3. RSQ - Milestone activity ordering violations

Milestone activities	a before b	$a = b$	a after b
a = DEPARTURE_READY b = ACTUAL_TIME_DEPART	434	66	0
a = ARRIVE_AT_RECEIVING_HOSPITAL b = DEPART_RECEIVING_HOSPITAL	470	29	1
a = ARRIVE_BACK_AT_BASE b = AVAILABLE_FOR_NEXT_TASKING	315	105	80

4.1   Preliminary Process Mining Analysis

In this section we complete the quality analysis by (i) generating event logs
from the sample respective data sets, and (ii) using PromLite 1.2 to perform
basic process discovery (Inductive Visual Miner plugin) and conformance analy-
sis (Multi-perspective Process Explorer plugin) to check that the event logs are
suitable for process mining.

4.1.1   QAS Process Discovery and Conformance

An event log was generated from the de-duplicated QAS sample data by (i)
treating each eARF in the sample data as a case, (ii) mapping the N_EARF
column to the event log case identifier attribute, (ii) creating an event from each

datetime column in the dataset by mapping the column name to the activity label and the row value of the column to the timestamp value. A process model was discovered and conformance checking (see Fig. 4) showed the model had 94.3% fitness (490 wrong and missing events out of 5,595 events in total). The discovered model highlighted some variations from the expected process behaviour as described by the QAS domain expert (illustrated in Fig. 2 and also highlighted some of the data quality issues discussed in 4.0.1. For instance, the expected behaviour is sequential execution of milestone tasks while the discovered model shows parallelism, (e.g. D_EN_ROUTE_VACIS and D_AT_SCENE_VACIS occur in a parallel block). Investigation showed that while there no cases where D_EN_ROUTE_VACIS preceded D_AT_SCENE_VACIS, there were 14 cases where the timestamp values for these two activities were the same. As observed earlier, the data quality analysis precision metric for the _VACIS times indicated only minute-level granularity. This may represent a "field dispatch" (i.e. non-tasked ambulance encounters an accident and notifies EMD it is on-scene). As such, the milestone events actually occurred in the expected sequence, but very close together (i.e. within the same minute) such that the recorded values were identical. In a similar vein, investigating the parallelism exhibited around the D_ON_SCENE_CAD and D_AT_PAT_VACIS activities showed that for the 581 cases where both activities occurred, in 198 cases the D_AT_PAT_VACIS activity occurred before the D_ON_SCENE_CAD activity. However, 174 of these cases had timestamps within 1 min of each other. Taking into account the minute-level granularity of the _VACIS times, it is again possible that these milestone events, in reality, occurred in the expected sequence, but very close together (i.e. within the same minute) but that the mixed granularity of the _VACIS and _CAD times results in incorrect event ordering. (We note that there were in fact 24 cases where there was 'real' deviation from the expected event ordering.)
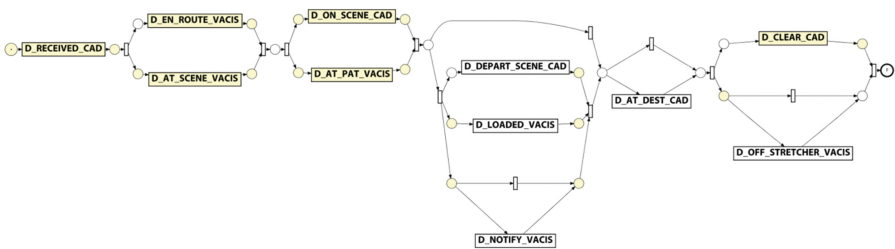

Fig. 4. QAS conformance model derived from sample data

Lastly, we note that the discovered model reflects the nature of the various types of attendance. For instance, (i) the 359 cases which skip the D_LOADED_VACIS and D_DEPART_SCENE_CAD steps reflect that not all attendances required the transport of a patient to hospital, and (ii) the 53 cases where a D_AT_DEST_CAD event occurs without a corresponding

D_OFF_STRETCHER_VACIS event which may reflect a non-transporting unit (e.g. Critical Care Paramedic backup) has proceeded to the hospital to accompany the transporting unit.

4.1.2   RSQ Process Discovery and Conformance

An event log was generated from the RSQ sample data by (i) treating each row in the sample data as an individual case, (ii) mapping the SOURCE_ID column to the event log case identifier attribute, and (iii) creating an event from each datetime column in the dataset by mapping the column name to the activity label and the row value of the column to the timestamp value. A process model was discovered and conformance checking (see Fig. 5) showed the model had 98.7% fitness (156 wrong and missing events out of 5,922 events in total). The discovered model highlighted some variations from the expected process behaviour as described by the RSQ domain expert (illustrated in Fig. 2 and also highlighted some of the data quality issues discussed in 4.0.2. The model shows parallelism for activities following AT_SCENE_PATIENT where the expected behaviour is sequential. The data quality assessment (see Table 3) and the model identified the activities and the extent of the deviation from expected behaviour. The conformance analysis revealed other event ordering issues including 10 cases where the first activity was not DATE_RETRIEVAL_REQUESTED.

5   Discussion and Lessons Learned

Data modelling prior to process mining informs the data extraction phase of the case study. The data models and relationship cardinalities show there are many possible *case* perspectives that are relevant (i.e. an incident may be considered a case, an individual patient experience may be considered a case, an individual response unit's dispatch/attendance/transport may be considered a case, etc.).
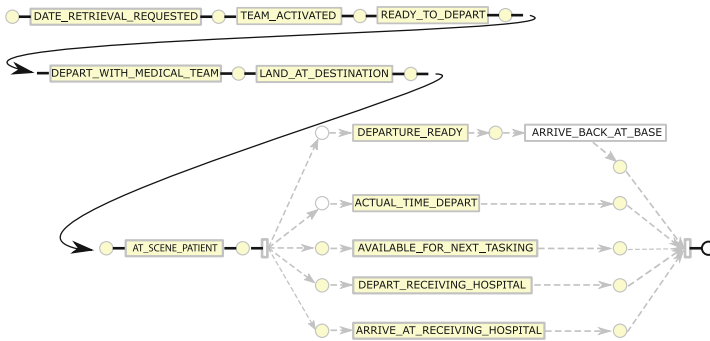

Fig. 5. RSQ conformance model derived from sample data

An important consideration in extracting the final dataset will be ensuring that, as well as the stakeholder's view that the patient experience is the case perspective, it will be possible to investigate other case perspectives.

The quality assessment of the (sample) data, conducted **prior** to the discovery and conformance analyses, adds value to the overall process mining exercise in at least **four** ways.

1. **Identifying event-data quality issues allows for the anticipation of certain observable features in subsequent process mining analysis.** For instance, the mixed granularity in timestamps led the analysts to anticipate incorrect event ordering (subsequently confirmed in the parallelism apparent in the discovered models). Further, the fact that the (RSQ) DATE_RETRIEVAL_REQUESTED values are all at day-level granularity precludes the possibility to properly assess performance aspects of various phases of aero-medical retrieval (for instance, how long does it take to activate a medical team following a retrieval request?). For the ground-based retrieval/transport data, the quality analysis showed duplication in the FIRST_ASSIGNED_CAD values. After discussion with QAS it emerged that it is QAS practice to include, for all response units dispatched to attend an incident, the same value for FIRST_ASSIGNED_CAD. This can be taken into account by making this a case attribute. Identifying this issue through quality assessment headed-off issues that may have arisen in the process mining analysis had the FIRST_ASSIGNED_CAD milestone been included as an activity for all eARFs and response units involved in the incident.

2. **Quantifying quality issues means that it is possible to separate systemic from occasional quality 'breaches'.** For instance, the fact that all (QAS) _VACIS timestamps were at a low level of precision (i.e. minute-level granularity) points to a systemic cause.

3. **Identifying quality issues allows for reasoning about the mechanisms that may have caused the event data quality issue.** For instance, it is unlikely that **all** (QAS) _VACIS events happened exactly on the minute, but, it is likely that, either the system recording the event had only minute-level precision, or that in extracting the data for analysis, seconds and milli-seconds were 'masked'. The fact that some (RSQ) cases have ARRIVE_AT_RECEIVING_HOSPITAL and DEPART_RECEIVING_HOSPITAL occurring at the same times may indicate a combination of human and system issues, i.e a human omission to record the ARRIVE time when the aircraft arrives (possibly due to patient care needs), and a system requirement that an ARRIVE time needs to be entered before a DEPART time can be entered.

4. **An understanding of 2 and 3 above facilitates informed engagement with process stakeholders and decisions about data quality remediation actions.** For instance, if the _VACIS granularity issues were as a result of incorrect data extraction, this quality issue can be resolved by simply extracting the data at the appropriate granularity.

Limitations associated with this current work include the fact that the approach has been trialled on only two, small data sets. Future work will focus on applications to larger datasets.

References

1. van der Aalst, W.M.: Extracting event data from databases to unleash process mining. In: vom Brocke, J., Schmiedel, T. (eds.) BPM - Driving Innovation in a Digital World. MP, pp. 105–128. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14430-6_8
2. Andrews, R., Suriadi, S., Wynn, M., ter Hofstede, A.H.: Healthcare process analysis. In: Process Modelling and Management for HealthCare. CRC Press (2017)
3. Badakhshan, P., Alibabaei, A.: Using process mining for process analysis improvement in pre-hospital emergency. In: Middle East North Africa Conference for Information Systems, Paris, March 2018 (2018)
4. Batini, C., Cappiello, C., et al.: Methodologies for data quality assessment and improvement. ACM Comput. Surv. (CSUR) **41**(3), 16 (2009)
5. Bose, R.J.C., Mans, R.S., van der Aalst, W.M.: Wanna improve process mining results? In: CIDM 2013, pp. 127–134 (2013)
6. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P.: PM^2: a process mining project methodology. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds.) CAiSE 2015. LNCS, vol. 9097, pp. 297–313. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19069-3_19
7. Halpin, T., Morgan, T.: Information Modeling and Relational Databases. Morgan Kaufmann, San Francisco (2010)
8. Lamine, E., Fontanili, F., Di Mascolo, M., Pingaud, H.: Improving the management of an emergency call service by combining process mining and discrete event simulation approaches. In: Camarinha-Matos, L.M., Bénaben, F., Picard, W. (eds.) PRO-VE 2015. IAICT, vol. 463, pp. 535–546. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24141-8_50
9. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B., Moleman, A.J.: Process mining in healthcare: data challenges when answering frequently posed questions. In: Lenz, R., Miksch, S., Peleg, M., Reichert, M., Riaño, D., ten Teije, A. (eds.) KR4HC/ProHealth 2012. LNCS (LNAI), vol. 7738, pp. 140–153. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36438-9_10
10. Rebuge, Á., Ferreira, D.R.: Business process analysis in healthcare environments: a methodology based on process mining. Inf. Syst. **37**(2), 99–116 (2012)
11. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. J. Biomed. Inform. **61**, 224–236 (2016)
12. Suriadi, S., Andrews, R., ter Hofstede, A., Wynn, M.: Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Inf. Syst. **64**, 132–150 (2017)

13. Wand, Y., Wang, R.: Anchoring data quality dimensions in ontological foundations. Commun. ACM **39**(11), 86–95 (1996)
14. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996)
15. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: PAKDDM, pp. 29–39 (2000)