

# *Ein Process-Mining und Machine-Learning basierter Ansatz zur Analyse von Prozessen*

*Stefan Peil*

## *Abstract*

*Das Gesundheitswesen ist aufgrund der steigenden Nachfrage nach medizinischen Leistungen, aber bei gleichzeitig erhöhtem Effizienzdruck eine anspruchsvolle Managementaufgabe. Es wurden in der vergangenen Jahren verschiedene Forschungsarbeiten durchgeführt, um klinische Prozesse zu optimieren, z. B. die Verkürzung der Wartezeit für die Konsultation, die Optimierung von Reservierungssystemen usw. In diesem Artikel schlagen wir eine Methode zur Analyse von klinischen Prozessen (Pfaden) vor, die auf der Methodik des Process Mining basiert. Process Mining zielt darauf ab, sachkundige Informationen aus in Informationssystemen aufgezeichneten Ereignisprotokollen zu extrahieren. Die vorgeschlagene Methodik im o.g. Projekt umfasst Datenintegration, Datenexploration, Datenanalyse und umfassende Diskussionsschritte mit den Prozessbeteiligten.*

**Keywords:** *Process Mining, Business Process Management, Data Mining, Markov-Prozesse, Warteschlangentheorie*

## **Einführung**

Der Vorteil des Process Mining gegenüber traditionellen Business Process Management Methoden besteht darin, automatisiert Prozessdaten zu extrahieren und in relevante Informationen umzuwandeln. Dies geschieht durch das Auslesen und der Analyse von Ereignisprotokollen, die in Informationssystemen gespeichert sind. Die Ereignisprotokolle für die Analyse werden in der Regel aus PAIS-Systemen (Process Awareness Information Systemen) wie Workflow-Management-Systemen, Datenbanksystemen, ERP -Systemen oder einem Krankenhausinformationssystem (KIS) erstellt. Nachfolgend ist die theoretische Vorgehensweise des Process Mining, die Architektur der beteiligten IT-Systeme sowie der Analyse- und Interpretationspfad von Prozessverbesserungen an einem Beispiel dargestellt.

## **Die Aspekte des Process Mining**

Grundsätzlich konzentriert sich Process Mining auf die Aspekte Discovery-, Conformance- und Enhancement von Prozessen. Nach einer eingehenden Analyse (Discovery) kann die Entscheidung getroffen werden, einen Prozess durch entsprechende Maßnahmen aus der Informationstechnologie zu bestätigen (Conformance) und im Managementsystem der Organisation festzuschreiben oder den Prozess anhand von im weiteren Verlauf dieser Arbeit beschriebener Kriterien zu optimieren (Enhancement).

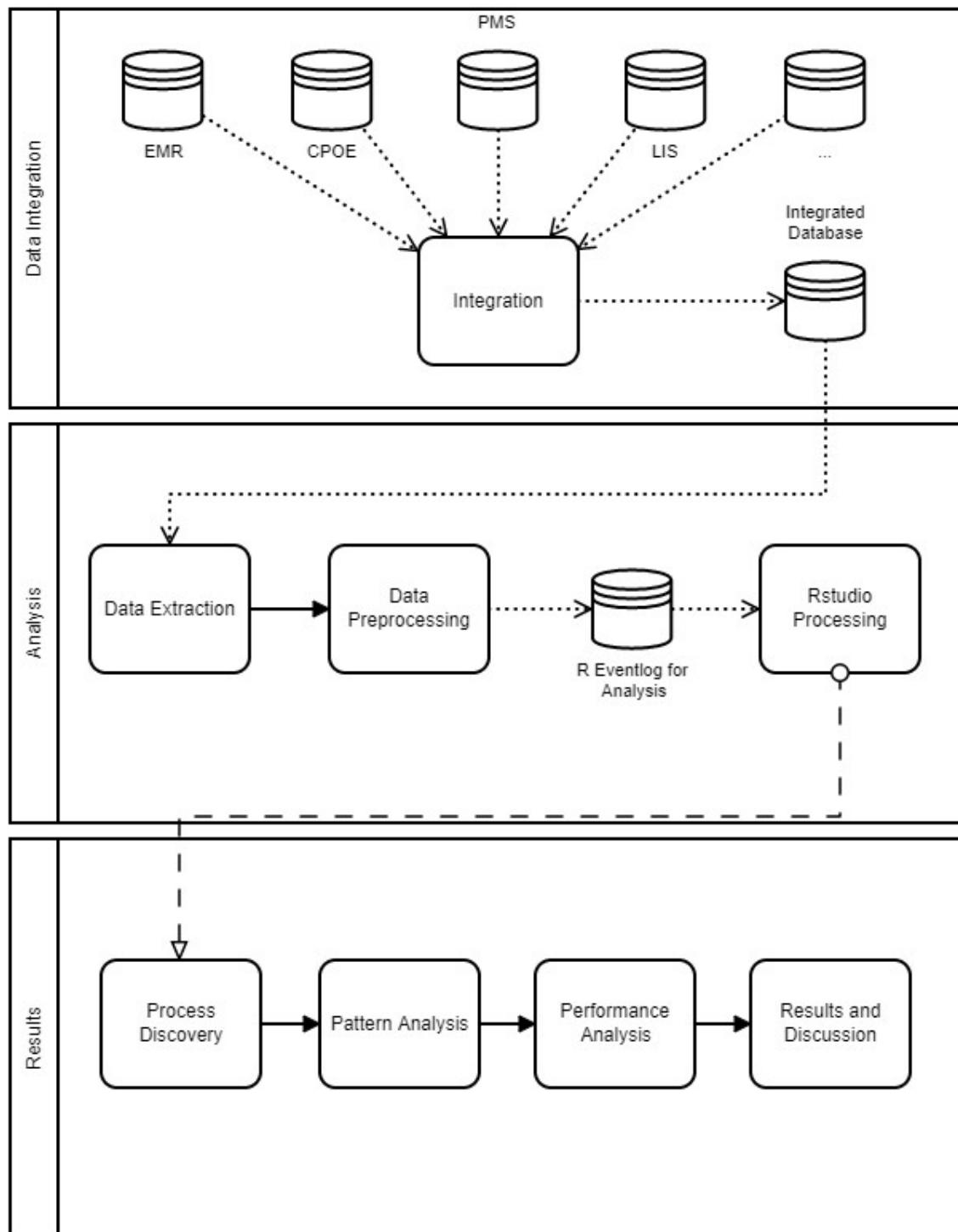
## **Das Process Mining Modell**

Im ersten Schritt wird hierzu der tatsächliche Prozessablauf aus den Ereignisdaten (Eventlog) erfasst und analysiert. Bei der Konformitätsprüfung wird das erfasste Prozessmodell anschließend mit dem beobachteten und dem explizit dokumentierte Soll-Verhalten verglichen. Anschließend besteht Verbesserung häufig darin, das erfasste Prozessmodell unter Verwendung von Attributen wie Zeit,

Kosten und durch Verringerung der Prozesskomplexität (Anzahl Schritte, ~Schnittstellen, ~Schleifen, Toolbrüche usw.) zu optimieren.

### Die Architektur und Systemlandschaft der Methode

Nachfolgend sind die Verknüpfungen der Systeme (Datenbanken) und der Analysepfad dargestellt. Der Analysepfad stellt den Weg von den Rohdaten (Data Integration) über die Analyse (Analysis) bis hin zu aussagekräftigen Informationen (Results) dar.



## Anwendung der Statistikprogrammiersprache R und der Anwendung RStudio

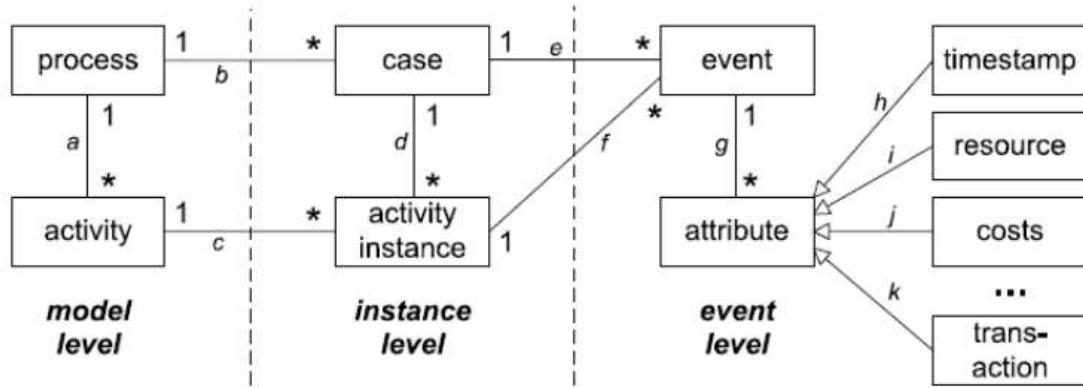
Im Rahmen dieses Projekts wird insbesondere bupaR [1] genutzt, eine Sammlung von R-Paketen, die für das Exploring und die Visualisierung von Ereignisdaten von Überwachungsprozessen verwendet werden können. Das Paket bupaR sowie weitere Pakete für die Software R und Rstudio bauen eine Brücke zwischen der BPM-Community (Geschäftsprozessmanagement) und den breiter aufgestellten R- und Data-Science-Communities, die sich mit sehr modernen mathematischen Modellen für die Analyse von Daten, bis hin zu KI-Anwendungen beschäftigen.

Darüber hinaus wird der breite Einsatz von Process Mining gefördert, da R ein sehr weit verbreitetes Open Source Analysewerkzeug ist, dessen Anwendungsgebiete für die Analyse und Visualisierung von Daten und Datenmodellen sowie die R-Community weltweit immer noch stark wachsen.

### Das Event-Data-Model als Grundlage der Analysen

*Eine grundlegende Darstellung vom Prozessablauf bis hin zu den auswertbaren Daten, bzw. Prozessmetadaten*

The notion of event log in bupaR refers to a set of events which are recorded in the context of a process. For instance, suppose the process under consideration takes place in the emergency department of a hospital. A general representation of the data model is shown below. Firstly, each event belongs to a case. A case, in general is an instance of the process. In the emergency department example, a case would be a visit by a patient.



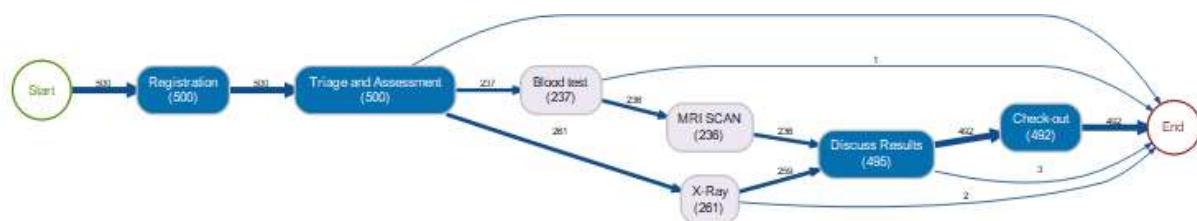
*Event data model (Quelle: [https://www.bupar.net/creating\\_eventlogs.html](https://www.bupar.net/creating_eventlogs.html))*

## Discovery, Conformance und Enhancement

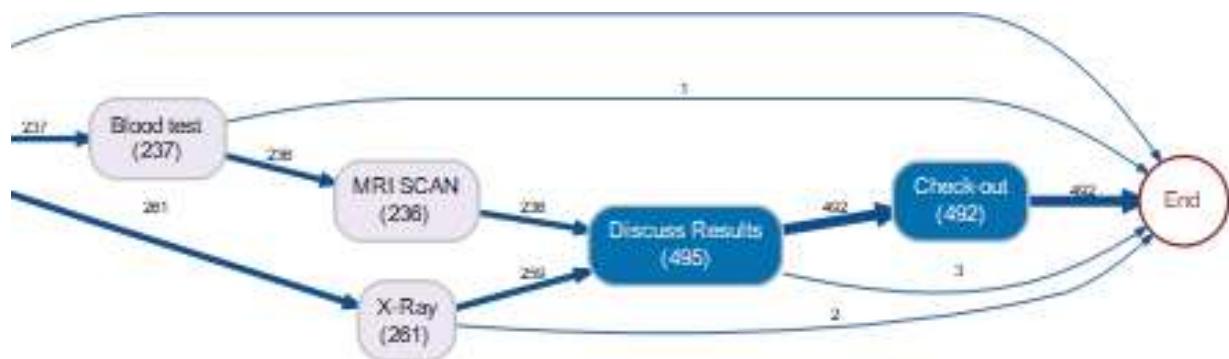
### (1) Process-Maps und Social-Networks

Aus dem Datensatz, dem aufbereiteten Eventlog, lassen sich Processmaps, also Darstellungen des Prozessablaufs erzeugen. Dies geschieht mit den oben dargestellten Daten von Case, timestamp, resource, activity usw.

Processmaps werden aus den Datensätzen errechnet. So lassen sich die tatsächlichen Abläufe ad hoc analysieren - ohne langwierige Befragung der Mitarbeiter. Optimierungspotenzial lässt sich sofort ermitteln. Die erzeugten Processmaps sind Grundlage von Soll-Ist-Vergleichen. Sie sind nicht nur für die personelle Organisation sondern auch für die betriebswirtschaftliche Analyse wichtig. Zudem sind die Processmaps auch im Rahmen des Qualitätsmanagements notwendige Werkzeuge zur Darstellung der Einhaltung von Standards und Pfaden.



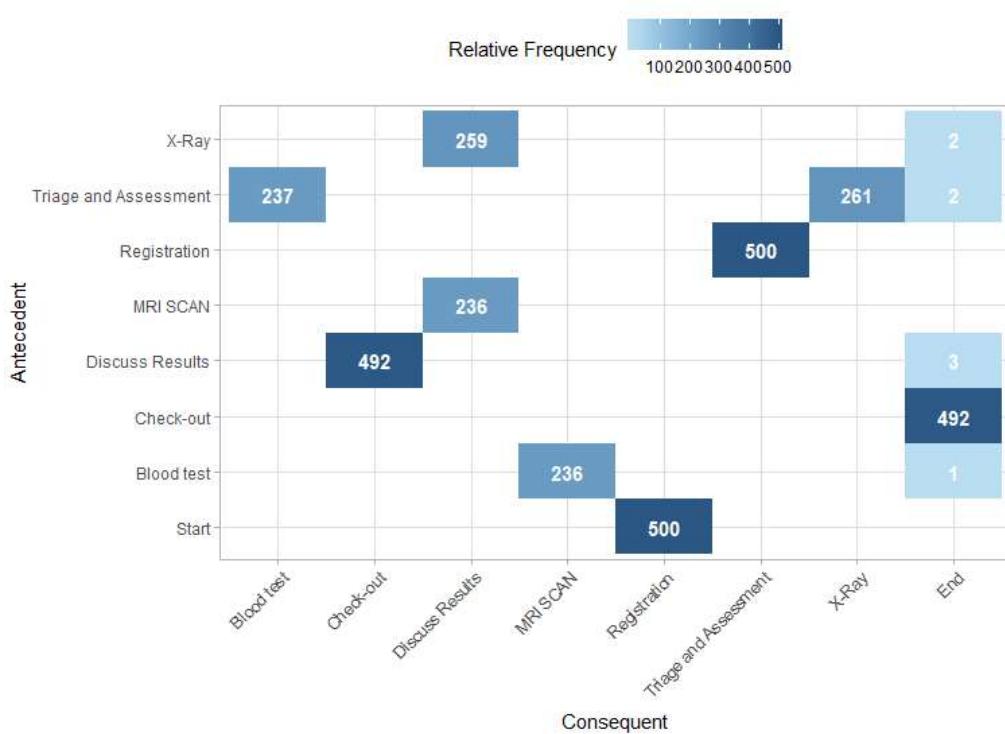
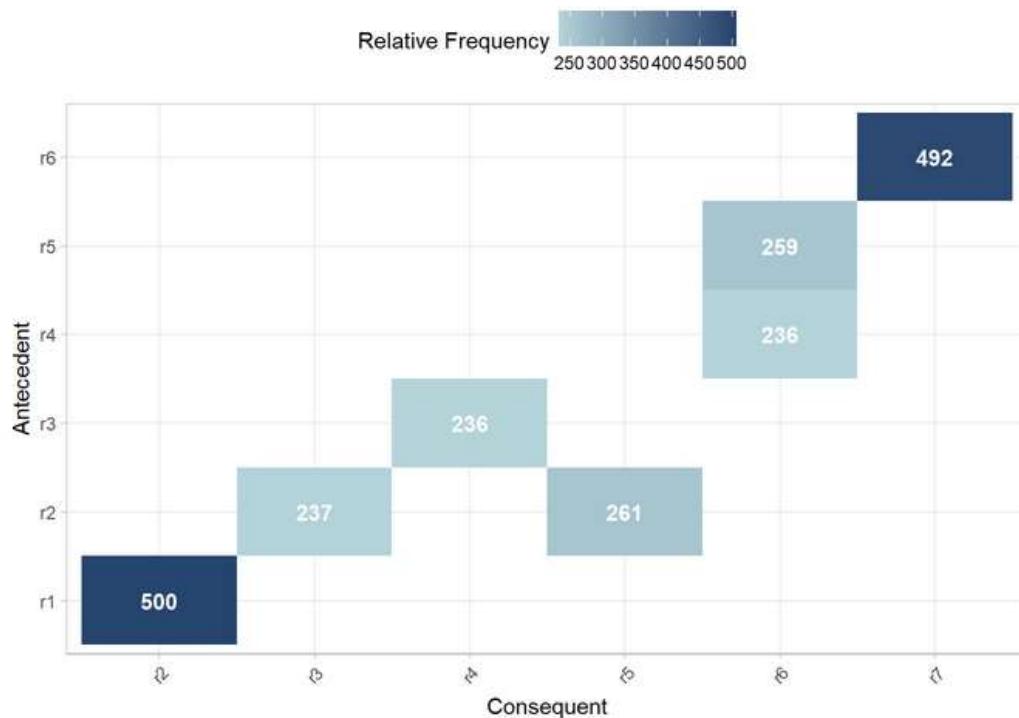
Ein Procesmap für die Organisation der Ablaufschritte



Ausschnitt aus dem oben dargestellten Ablauf (Anzahl der Cases)

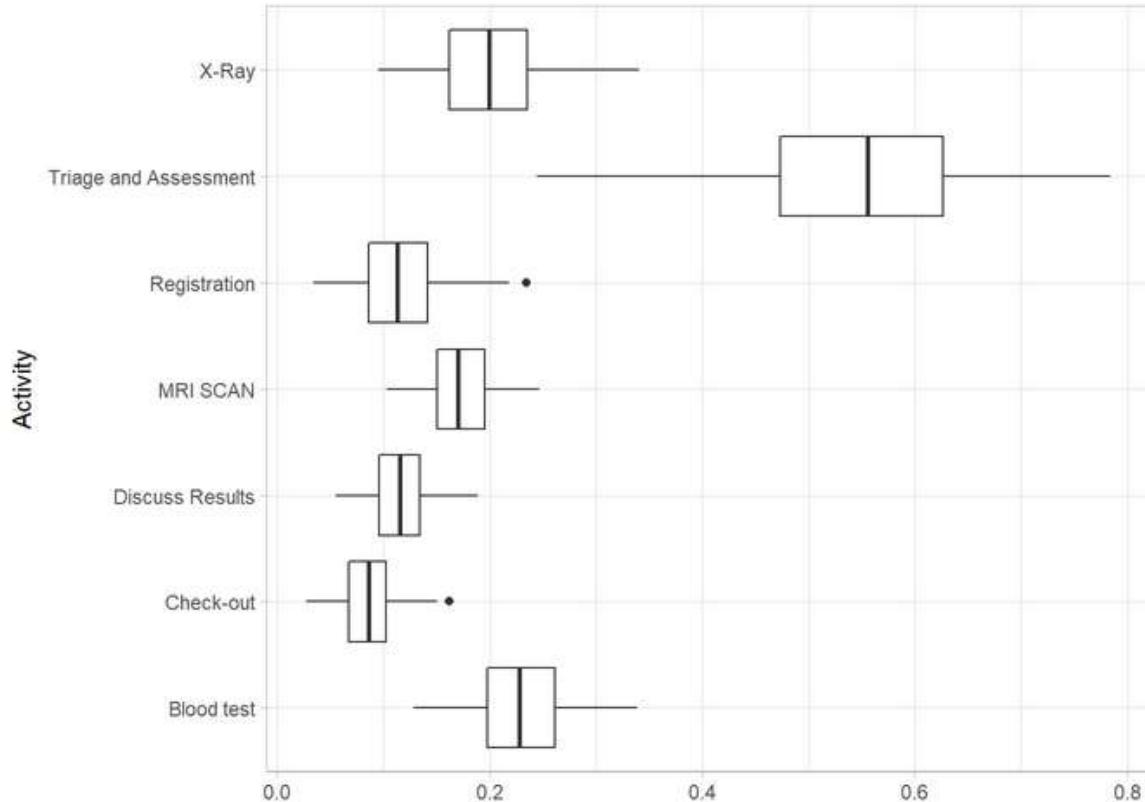
## (2) Precedent Matrix

Die Precedent Matrix zeigt die Anzahl der direkten Kontakte zwischen den Ressourcen oder Prozessschritten (Vorgänger Antecedent) und Nachfolger (Consequent) im ermittelten Eventlog an. Hiermit lässt sich die Auslastung einer personellen Resource oder eines Funktionsbereiches schnell bewerten.



### (3) Die Auswertung von Prozesszeiten

Die Prozesszeiten können auf verschiedenen Ebenen berechnet werden. Sie können nur berechnet werden, wenn Start- und Endzeitstempel für Aktivitätsinstanzen verfügbar sind. Die „Boxen“ lassen sich wie folgt erklären: Die linke Kante ist das 25% Quartil, der mittlere dicke Strich ist der Median, also das 50% Quartil und die rechte Kante ist das 75% Quartil. Die horizontalen Striche „Whisker“ zeigen vereinfacht ausgedrückt Ausreißer an. Mit den „Box-Whisker-Plots“ lassen sich sehr schnell die Prozesszeiten und deren Streuung für Prozesse erkennen.



Prozesszeiten in der Darstellung als Boxplot

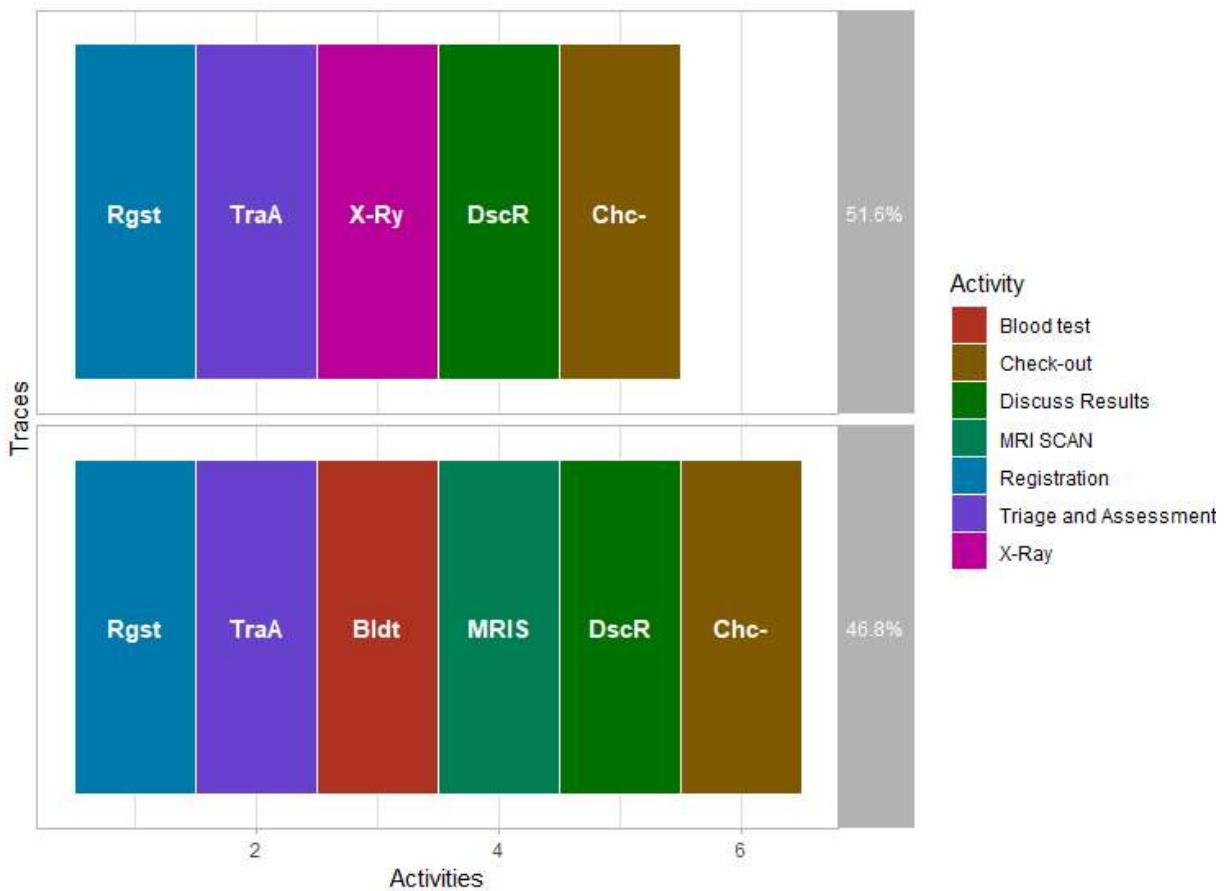
	handling	min	q1	mean	median	q3	max	st_dev	iqr	total	relative_frequency
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Registration	0.0345	0.0863	0.115	0.113	0.142	0.235	0.0397	0.0555	57.4	0.184
2	Triage and Assessment	0.245	0.473	0.546	0.556	0.626	0.783	0.115	0.153	273.	0.184
3	Discuss Results	0.0556	0.0964	0.116	0.116	0.134	0.189	0.0261	0.0377	57.3	0.182
4	Check-out	0.0278	0.0671	0.0860	0.0863	0.103	0.162	0.0258	0.0358	42.3	0.181
5	X-Ray	0.0956	0.162	0.202	0.200	0.235	0.340	0.0534	0.0734	52.7	0.0959
6	Blood test	0.129	0.198	0.231	0.228	0.261	0.339	0.0442	0.0630	54.7	0.0871
7	MRI SCAN	0.104	0.150	0.173	0.170	0.196	0.247	0.0306	0.0454	40.8	0.0867

Einzelergebnisse der Prozesszeiten zur weiteren Berechnung

Oben sind die Daten in einer Tafel dargestellt.

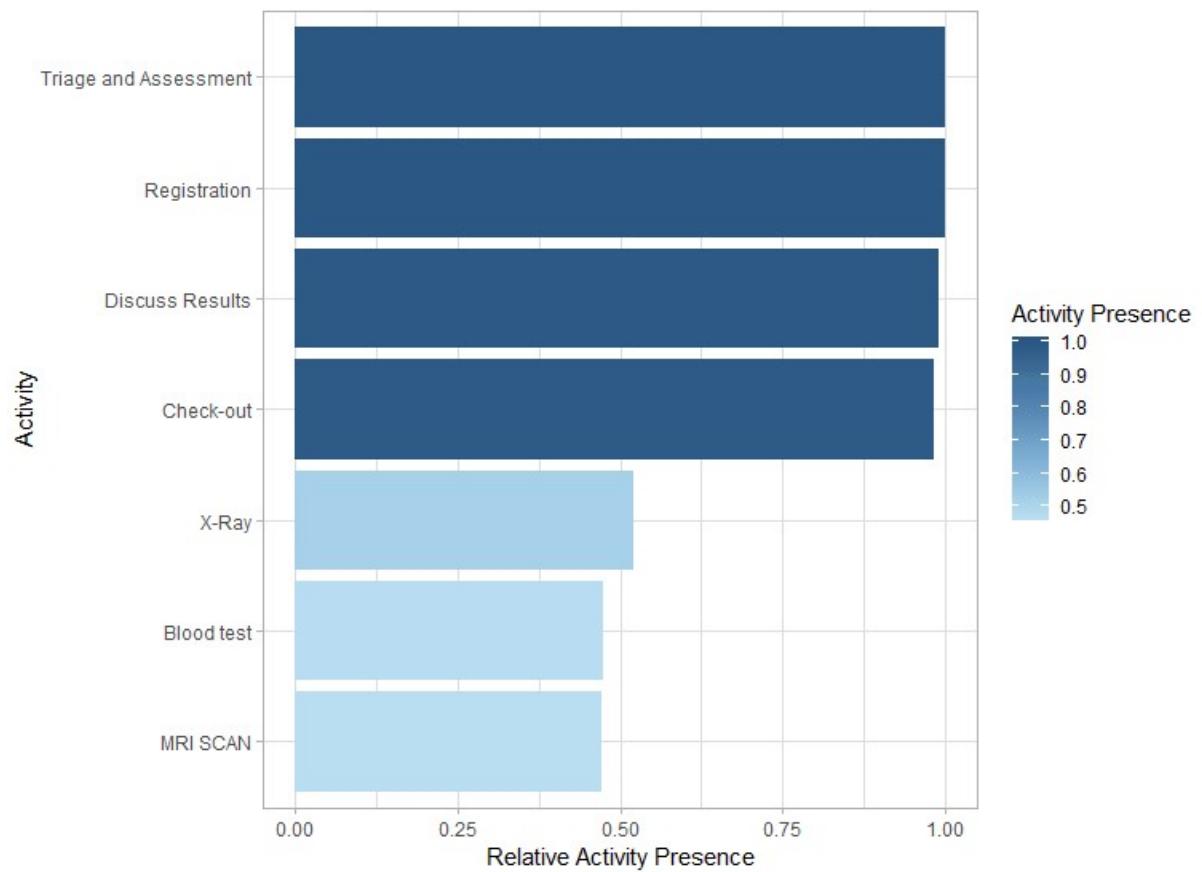
#### (4) Trace Explorer

Mit dem trace explorer können verschiedene Aktivitätsabläufe im Ereignisprotokoll visualisiert werden. Er kann verwendet werden, um häufige sowie seltene „Spuren“ zu untersuchen. Das Coverage-Argument gibt an, wie viel des Protokolls Sie untersuchen möchten. Standardmäßig ist der Wert auf 0,2 eingestellt, d.h. er zeigt die meisten (In-) Frequenzverläufe, die mindestens 20% des Ereignisprotokolls abdecken.



## (5) Activity Presence

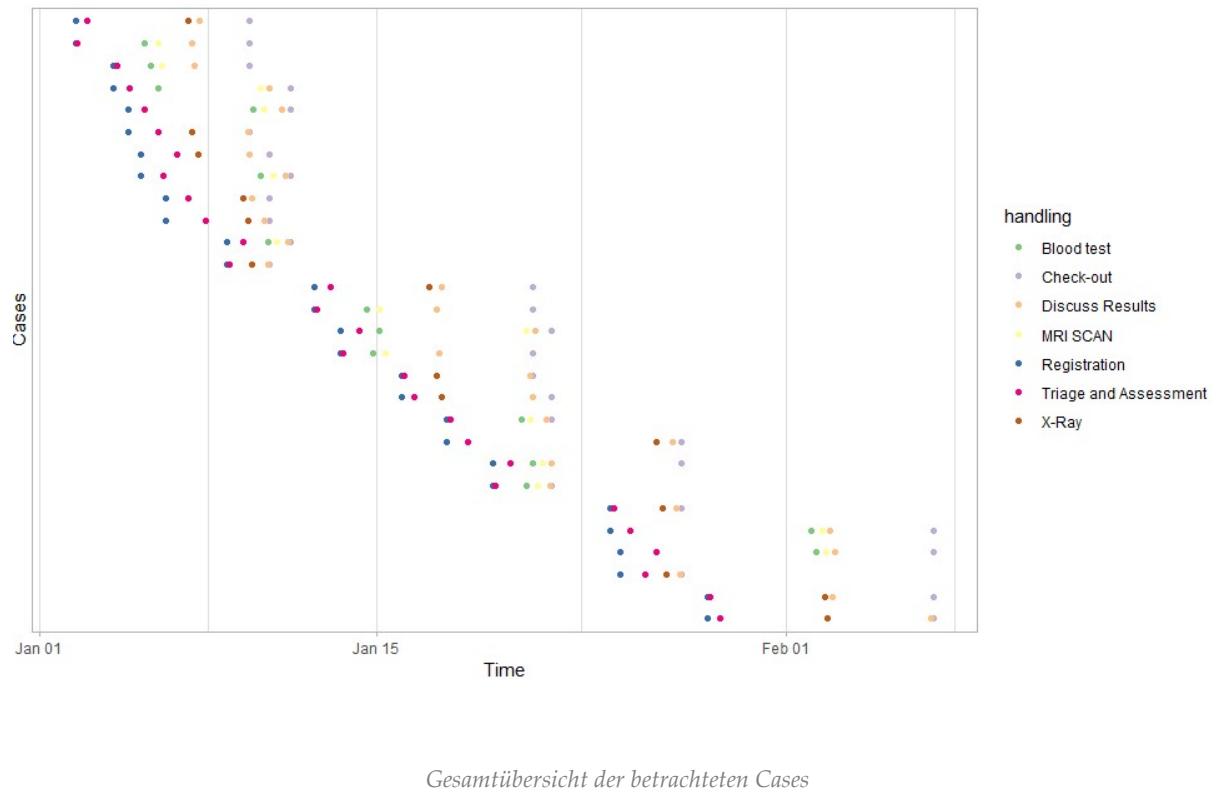
Die Präsenz von Aktivitäten zeigt, in wie viel Prozent eines Falles (Case) eine bestimmte Aktivität vorliegt. Es hat kein Level-Argument. Im Beispiel ist zu erkennen, dass jeder Case eine Registration und ein Triage und Assessment durchlaufen hat. Bei ca. 50% aller Patienten wurden Röntgen, Bluttest und MRT durchgeführt.



*Activity Presence*

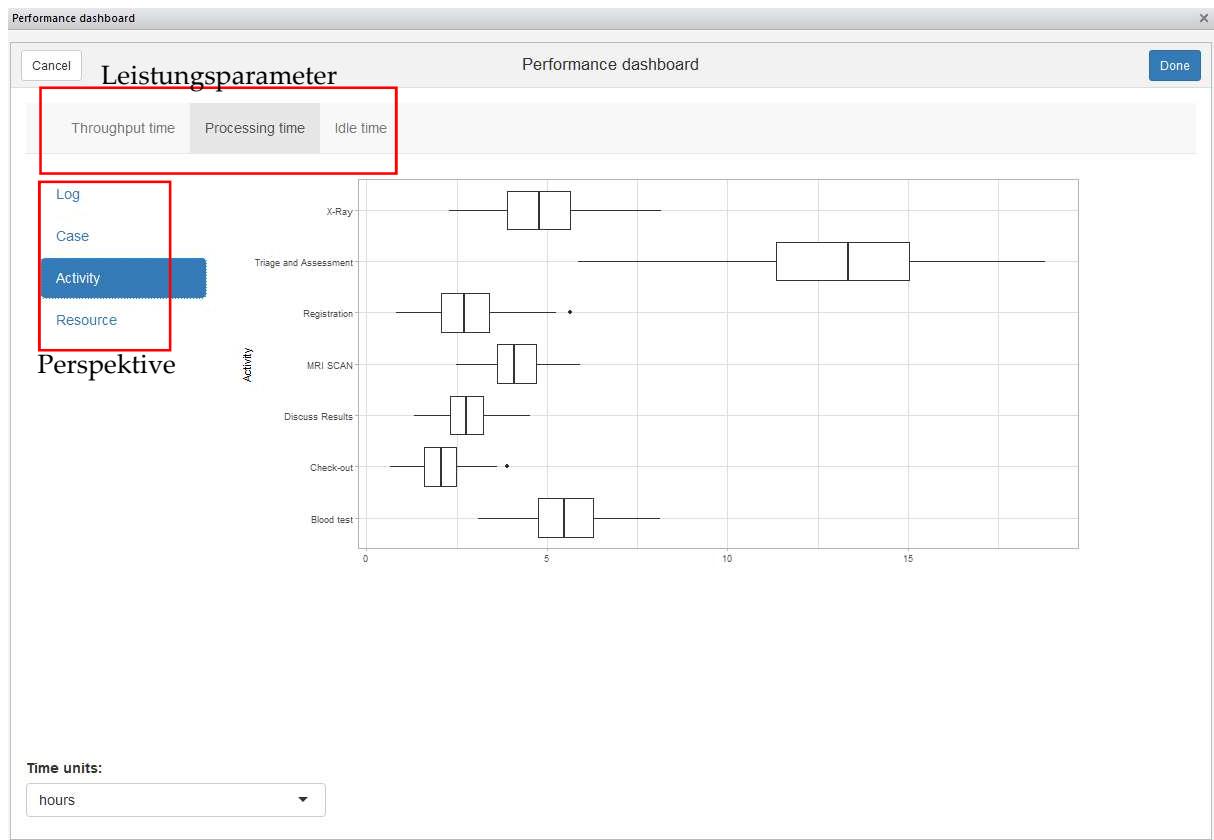
## (6) Gesamtübersicht

Zusammenfassen lassen sich die Prozesszeiten über einen frei definierbaren Zeitraum. So können Verbesserungen oder Potenziale sogar aus speziellen Fällen (Cases) identifiziert werden



## (7) Dashboards

Das processmonitR-Paket enthält mehrere vordefinierte Dashboards, um den Prozess basierend auf Ereignisprotokollen zu überwachen. Gibt es eine Routine für die Erzeugung von Eventlogs, können damit Auswertungen quasi in „Echtzeit“ erledigt werden.



Performannce Dashboard

## (8) Leistungsbewertung von Abläufen mit statistischen Modellen

Mit Hilfe der Warteschlangentheorie lassen sich Prozesse manuell sehr gut berechnen und angedachte Optimierungen vor einer Veränderung von Prozessen bewerten.

Untersuchungen des Leistungsverhaltens sind in verschiedenen Phasen eines Prozesslebenszyklus (Prozessentwurf, Konfiguration, Betrieb und Optimierung) notwendig. Ziel ist stets eine sinnvolle Abwägung zwischen Resourcenaufwand und -belastung und Qualität der erbrachten Leistung.

Prozessanalysen sind i. d. R. stochastisch oder operational. Stochastische Modelle liefern als Ergebnis ebenfalls stochastische Kenngrößen. Operationale Analysen liefern gemessene Kenngrößen. Der Vergleich beider Kenngrößen ist im Rahmen der Optimierung von Prozessen eine hervorragende Methode um Potenziale und auch Entwurfsfehler aufzudecken.

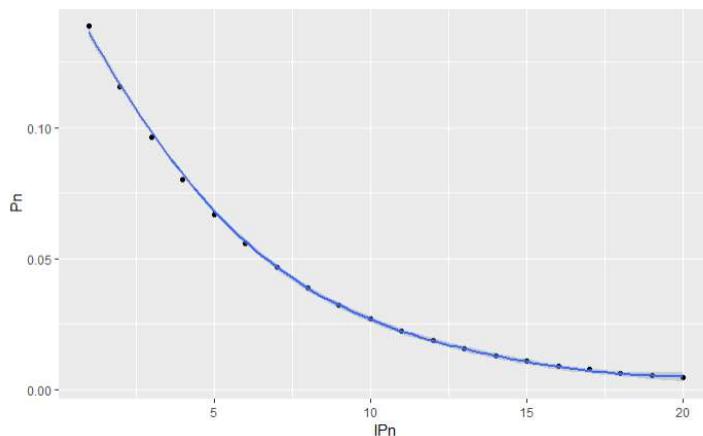
Bei der Analyse eines Notfallprozesses (Triage) lassen sich beispielsweise folgende Kenngrößen ermitteln und Fragen beantworten. Dazu nachfolgend ein parametrisiertes Beispiel:

<b>lbda</b> <dbl>	<b>mu</b> <dbl>	<b>l</b> <dbl>	<b>ro</b> <dbl>
10	12	20	0.8333333

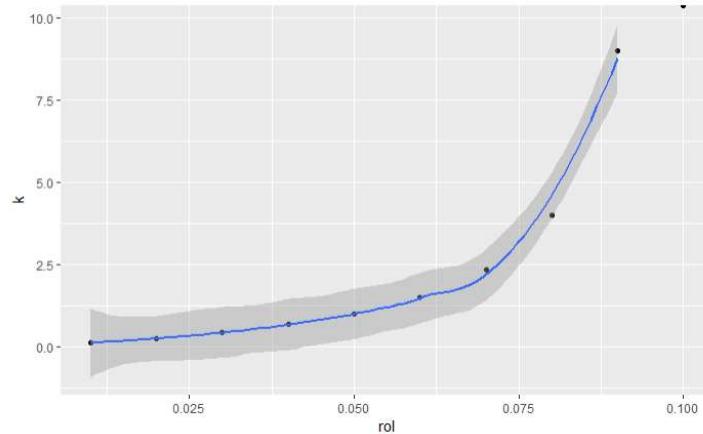
lbda = Ankunftsrate von Patienten in den Prozessschritt Triage, mu = Bedienrate (triage), l = Hilfsgröße (Iterator), ro = Auslastungsgrad im Prozessschritt (< 1)

Beispielhafte Berechnung von zwei Kenngrößen:

1. Wahrscheinlichkeit, dass sich eine Anzahl LPn (1..20) Patienten im Prozess Triage befinden:



2. Mittlere Anzahl k erwarteter Patienten im Prozessschritt Triage unter Berücksichtigung der Auslastung rol (rol 0 .. < 1) im Prozessschritt:



Für den oben beschriebenen Prozessschritt Triage lassen sich noch weitere Kenngrößen berechnen:

Erwartete Anzahl der Patienten im System (Warten und Trage) = 5 Patienten

Erwartete Anzahl Patienten in der Warteschlange = 4 Patienten

Mittlere Zeit für das Warten und Triage = 0.5 h (ohne Streuung)

Mittlere Zeit für Triage = 0.1 h (ohne Streuung)

Die oben berechneten Kenngrößen werden mit gemessenen Werten verglichen, adjustiert und können Bestandteil des Qualitätsberichts sowie Grundlage von Mittelanforderungen an die Geschäftsleitung sein.

[1] Zitation bupaR:

<https://www.sciencedirect.com/science/article/pii/S0950705118305045> Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., & Vanhoof, K. (2019). bupaR: Enabling reproducible business process analysis. *Knowledge-Based Systems*, 163, 927-930.

sciencedirect.com

bupaR: Enabling reproducible business process analysis

---

## Glossar

EMR: Electronical Medical Records

CPOE: Computerized Physician Order Entry

PMS: Patientverwaltung

LIS: Laborinformationssystem