

MESTRADO EM
MÉTODOS QUANTITATIVOS E DECISÃO ECONÓMICA
E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
RELATÓRIO DE ESTÁGIO

ANÁLISE DE CONFORMIDADE EM PROCESSOS DE SAÚDE:
ESTUDO DO CANCRO COLORRETAL NO HOSPITAL DA LUZ DE
LISBOA

HUGO MIGUEL TEIXEIRA PINTO FERREIRA PACHECO

Janeiro - 2020

MESTRADO EM
MÉTODOS QUANTITATIVOS E DECISÃO ECONÓMICA
E EMPRESARIAL

ANÁLISE DE CONFORMIDADE EM PROCESSOS DE SAÚDE:
ESTUDO DO CANCRO COLORRETAL NO HOSPITAL DA LUZ DE
LISBOA

HUGO MIGUEL TEIXEIRA PINTO FERREIRA PACHECO

ORIENTAÇÃO:

PROFESSORA DOUTORA TERESA MELO (SAARLAND UNIVERSITY OF
APPLIED SCIENCES, SAARBRÜCKEN)

DOUTOR FILIPE COSTA (HOSPITAL DA LUZ, LISBOA)

Janeiro - 2020

Agradecimentos

Esta etapa nunca teria sido ultrapassável sem o apoio de certas pessoas a quem deixo aqui o meu agradecimento. Quero começar por agradecer aos meus orientadores Dr. Filipe Costa do Hospital da Luz de Lisboa e à Prof. Teresa Melo da HTW Saar que me apoiaram durante este período e que responderam a todas as minhas dúvidas e questões. Deixo um enorme obrigado a toda a equipa do Hospital da Luz - Learning Health por tudo terem feito para me integrar e me apoiar no meu local de estágio, foram todos absolutamente incríveis. Aos meus amigos que não só durante esta fase, mas durante toda a minha vida me deram todo o apoio e tornaram tudo mais fácil. Quero deixar também um agradecimento especial ao Fábio Marques e ao Tiago Duarte que não só me guiaram durante o meu período no ISEG, dentro e fora das aulas, mas foram também a influência que me fez seguir este mestrado. Por último quero dar o maior obrigado de todos à minha família, em especial aos meus pais e irmã que me apoiaram incondicionalmente e me deram as condições para atingir todos os meus objetivos académicos e pessoais.

Resumo

Este relatório foca a descrição da metodologia usada para executar a análise de conformidade em processos de saúde para o Cancro Colorrectal no Hospital da Luz de Lisboa. Esta metodologia foi desenvolvida com o objetivo de potenciar uma compreensão mais eficiente do processo de tratamento a que cada paciente é sujeito nesta instituição e, principalmente, para estimar a conformidade entre processos de tratamento individuais (entre 2017 e 2018) e o processo de tratamento padrão atualmente aplicado na instituição supracitada. Adicionalmente, pretende relacionar o nível de conformidade e o custo inerente ao tratamento de cada doente para avaliar o impacto da conformidade na eficiência de custos. Os custos apresentados não são reais devido a legislação relativa à proteção de dados, contudo as relações entre os dados espelham a realidade.

Este estudo surge no seguimento de anteriores neste campo promovidos pelo Hospital da Luz de Lisboa, para apoiar uma futura aplicação da teoria de *Value Based Healthcare*. Pretende-se com esta abordagem alcançar não só uma melhoria dos serviços prestados, mas também dos resultados clínicos dos utentes aliado a custos mais justos para ambas as partes.

Palavras chave: Process mining, Análise de conformidade, Saúde, Cancro Colorrectal.

Abstract

This final report aims at describing the methodology used to perform a conformance analysis in the healthcare process for the colorectal cancer in the Hospital da Luz de Lisboa. This methodology was developed enable a better understanding of the treatment process a patient undergoes in this institution and mainly to assess the conformity of the individual treatment process (in 2017 and 2018) and the standard treatment processes currently in place at the aforementioned institution. In addition, focus is also given to relate the conformity calculated and the costs related to each patient, in order to evaluate the impact of the conformity in the cost efficiency. The costs provided in the study are not real due to the data protection legislation, nevertheless the relations between the data mirrors reality.

This is yet another study on this field promoted by the Hospital da Luz de Lisboa to support a future establishment of the Value based Healthcare. The aim is to improve not only the services, but also the outcomes of patients and to make the pricing fairer to the patient and to the institution.

Key words: Process mining, Conformance checking, Healthcare, Colorectal Cancer.

Índice

Agradecimentos	III
Resumo	IV
Abstract.....	V
Índice.....	VI
Índice de Ilustrações	VIII
Índice de Tabelas	VIII
Índice de Figuras	VIII
Lista de Siglas	IX
CAPÍTULO 1: INTRODUÇÃO.....	10
1.1 Objetivo do estágio	10
1.2 Descrição do problema	11
1.3 Estrutura do relatório de estágio	12
CAPÍTULO 2: ENQUADRAMENTO TEÓRICO	13
2.1 Cancro colorrectal	13
2.2 Clinical Pathway	14
2.3 Process Modeling.....	15
2.4 Event Log.....	19
2.5 Process Mining.....	20
2.6 Análise de Conformidade.....	22
2.7 ProM.....	23
2.8 Técnicas alternativas de análise de conformidade	24
CAPÍTULO 3: METODOLOGIA	26
3.1 Técnica de pesquisa	26
3.2 Dados	26
3.3 Processamento dos dados	27
3.4 Modelação processual do Clinical Pathway.....	28
CAPÍTULO 4: RESULTADOS.....	33
CAPÍTULO 5: DISCUSSÃO.....	39

CAPÍTULO 6: CONCLUSÕES	40
Bibliografia	43
ANEXO 1 – TABELA DE DADOS E GRÁFICOS DE DISPERSÃO	45
ANEXO 2 – MODELO BPMN DO PROCESSO DE TRATAMENTO DO CCR SEGUNDO O HOSPITAL DA LUZ DE LISBOA.....	47
ANEXO 3 – MODELO BPMN PARA O TRATAMENTO DO CCR ADAPTADO	50

Índice de Ilustrações

ILUSTRAÇÃO 2.1 DIAGRAMA DO CLINICAL PATHWAY PARA O CANCRO DO RETO NO HOSPITAL DA LUZ LISBOA.....	5
ILUSTRAÇÃO 2.2 ELEMENTOS DA LINGUAGEM BPMN	6
ILUSTRAÇÃO 2.3 EXEMPLO DE DIAGRAMA BPMN COM LANES	7
ILUSTRAÇÃO 2.4 EXEMPLO DIAGRAMA BPMN COM ANNOTATIONS	8
ILUSTRAÇÃO 2.5 DIAGRAMA EXPLICATIVO DE PROCESS MINING	10
ILUSTRAÇÃO 3.1 CÓDIGO DE VBA USADO PARA PROCESSAMENTO DE DADOS.....	17
ILUSTRAÇÃO 4.1 GRÁFICO CIRCULAR COM O PESO EM PORCENTAGEM DO CUSTO DE CADA ATIVIDADE EM RELAÇÃO AO CUSTO TOTAL.....	24
ILUSTRAÇÃO 4.2 BOX PLOT DA TRACE FITNESS	25
ILUSTRAÇÃO 4.3 DIAGRAMA DE DISPERSÃO ENTRE O TRACE FITNESS E O NÚMERO DE ATIVIDADES	26
ILUSTRAÇÃO 4.4 DIAGRAMA DE DISPERSÃO ENTRE O TRACE FITNESS E O CUSTO TOTAL.....	27
ILUSTRAÇÃO 4.5 DIAGRAMA DE DISPERSÃO ENTRE O TRACE FITNESS E O DESVIO ABSOLUTO EM RELAÇÃO AO CP	27

Índice de Tabelas

TABELA 1 EXEMPLO DE EVENT LOG FORNECIDO PELO HOSPITAL DA LUZ.....	9
---	---

Índice de Figuras

FIGURA 1 TABELA DE DADOS DE PACIENTES E ATIVIDADES	35
FIGURA 2 GRÁFICO DE DISPERSÃO ENTRE NÚMERO DE ATIVIDADES E CUSTO TOTAL	36
FIGURA 3 GRÁFICO DE DISPERSÃO ENTRE NÚMERO DE ATIVIDADES E DESVIO ABSOLUTO EM RELAÇÃO AO CP	36
FIGURA 4 PARTE 1 DO MODELO BPMN PARA O TRATAMENTO DO CCR SEGUNDO O HOSPITAL DA LUZ	37
FIGURA 5 PARTE 2 DO MODELO BPMN PARA O TRATAMENTO DO CCR SEGUNDO O HOSPITAL DA LUZ	39
FIGURA 6 PARTE 3 DO MODELO BPMN PARA O TRATAMENTO DO CCR SEGUNDO O HOSPITAL DA LUZ.....	39
FIGURA 7 MODELO BPMN PARA O TRATAMENTO DO CCR ADAPTADO	40

Lista de Siglas

API – Application Programming Interface

B2B – Business to business

BPM – Business Process Management

BPMN – Business Process Management Notation

CCR – Cancro Colorretal

CP – Clinical Pathway

CSV – Comma Separated Values

ERP – Enterprise Resource Planning

HIS – Hospital Information System

MCDT – Métodos Complementares de Diagnóstico e Tratamento

SI – Sistema de Informação

SWRL–Semantic Web Rule Language XES–eXtensible Event Stream

UML – Unified Modeling Language

CAPÍTULO 1: INTRODUÇÃO

O sector da saúde tem vindo a sentir uma pressão crescente para melhorar os seus processos de gestão e de prestação de serviços devido ao constante aumento da despesa em saúde. O agravamento da despesa em saúde é esperado em todas as áreas médico-cirúrgicas quer pela pressão do crescimento quer do envelhecimento populacional, sendo esta tendência expectável também no caso particular do Cancro Colorrectal (CCR). Dessa forma, a otimização dos processos de tratamento é essencial para conseguir, simultaneamente, melhorar os cuidados prestados aos utentes e controlar os custos associados a esses mesmos cuidados.

Neste contexto foram propostas duas questões para teste: (1) perceber se o *clinical pathway* (explicado no capítulo seguinte) definido para o tratamento do CCR é de facto seguido pelos profissionais de saúde, e (2) avaliar os custos associados aos processos de tratamento realizados.

O objetivo final é perceber se variações na adesão dos profissionais ao *clinical pathway* têm ou não impacto nos custos de prestação do serviço.

1.1 Objetivo do estágio

O presente relatório foi elaborado durante o estágio realizado no Hospital da Luz – Learning Health ao abrigo de um protocolo estabelecido entre o Hospital, o Instituto Superior de Economia e Gestão e o mestrando. A Learning Health, empresa pertencente ao grupo Luz Saúde, é uma empresa de formação, inovação e investigação em saúde. A sua atividade fomenta a realização de programas de formação e simulação para profissionais internos e externos (tanto de saúde como de outras áreas de interesse), de eventos científicos (simpósios e colóquios), de investigação suportada por métodos de analítica avançada e engenharia de fatores humanos, bem como de programas de aceleração (para *start-ups*). Pauta-se por um ambiente extremamente multidisciplinar e motivador que proporciona a quem tem a oportunidade de integrar a equipa a possibilidade de aprender muito sobre diferentes áreas relacionadas com o sector da saúde numa base constante.

O interesse fundamental do estágio que serviu de base a este relatório, foi a aplicação dos conhecimentos adquiridos ao longo do mestrado em Métodos Quantitativos e Decisão Económica e Empresarial,

leccionado no Instituto Superior de Economia e Gestão, aliando-se à exploração de numa nova área de conhecimento para mim, enquanto mestrando e futuro profissional.

Para o desenvolvimento deste estudo utilizou-se uma ferramenta de *process mining* (ProM tools) em conjunto com conhecimentos de estatística e avaliação financeira, aplicados na análise processual e de custo, com vista a otimizar o processo e apoiar, desta forma, a empresa na tomada de decisão quanto ao mesmo.

1.2 Descrição do problema

No contexto da crescente pressão de otimização processual sobre as instituições do sector da saúde, em particular no Hospital da Luz, devido ao aumento da despesa em métodos de diagnóstico e terapêutica e ao rápido desenvolvimento de sistemas de informação hospitalar, surge a necessidade de avaliar se o *clinical pathway* desenvolvido internamente para o tratamento do CCR é efetivamente adotado pelas equipas responsáveis.

Os registos hospitalares são tendencialmente mais completos e a sua análise torna-se, nessa dependência, mais complexa. Estes registos são clusters de informação de validade potencial extrema no âmbito da gestão empresarial, muito frequentemente inexplorada apesar de acessível, que deve, por isso, ser tratada de forma cuidada com métodos capazes de extrair ao máximo as suas valências escondidas. Com este intuito, foram aplicados vários métodos aos dados referentes a 2017 e 2018 sobre os tratamentos ao CCR dos utentes do Hospital da Luz para (1) tentar perceber se os procedimentos efetuados coincidem com os procedimentos pré-estabelecidos, e (2) averiguar se a variação teve algum impacto sobre a eficiência de custo desses mesmos processos. Utilizaram-se técnicas de *process mining* (definido no próximo capítulo) para a verificação de conformidade e uma direta comparação de custos reais e custos teóricos para avaliar o efeito da supracitada variação.

1.3 Estrutura do relatório de estágio

O presente relatório está dividido em sete capítulos.

No presente capítulo, define-se o objetivo do estágio e faz-se a descrição do problema em que este vai incidir.

O segundo capítulo trata o enquadramento teórico onde se explora o estado da arte e explanam os conceitos essenciais para a contextualização da problemática abordada no corrente documento, começando pela definição da patologia de interesse, a descrição do *clinical pathway*, ou seja, do processo estandardizado preconizado pela instituição e transitando de seguida para os conceitos mais relacionados com a disciplina utilizada (*process mining*) no processamento dos dados, com a apresentação do *process modeling*, da análise de conformidade, do *event log* e do software de *process mining* utilizado.

No terceiro capítulo, é abordada de forma mais detalhada a metodologia utilizada no desenvolvimento do trabalho para a obtenção de informação e resultados, desde a pesquisa bibliográfica até à aplicação do algoritmo utilizado para a análise de conformidade.

No quarto capítulo, caracterizam-se de forma sucinta os resultados mais relevantes para o estudo.

No quinto capítulo, detalha-se a discussão dos resultados.

No sexto capítulo, destacam-se as conclusões, discutidas numa perspetiva de análise dos resultados obtidos.

No sétimo e último capítulo, pretende-se sublinhar a importância de trabalhos subsequentes nesta linha de investigação, tendo em conta o potencial da aplicabilidade destas conclusões no sector empresarial.

CAPÍTULO 2: ENQUADRAMENTO TEÓRICO

Neste capítulo introduzem-se os conceitos essenciais para que o leitor menos familiarizado com a área de estudo consiga ter compreensão do que é apresentado.

No ponto 2.1 faz-se uma breve apresentação da patologia, sendo referido o processo de tratamento. Em 2.2 é definido o conceito de *clinical pathway* e as suas principais características. Em seguida, no ponto 2.3, explica-se em que consiste o *process modeling* e é apresentada também a linguagem selecionada para executar essa mesma modelação neste estudo. No ponto 2.4 descreve-se o que é um *event log* e a sua utilidade para este estudo. Adicionalmente, no ponto 2.5, apresenta-se o conceito de *process mining* no geral e mais em detalhe a análise de conformidade no ponto seguinte (2.6) já que será a vertente de *process mining* usada nesta análise. No ponto 2.7, especifica-se o software usado para a aplicação de *process mining*. Por último, no ponto 2.8 apresentam-se algumas técnicas de análise de conformidade como possíveis alternativas à qual foi utilizada.

2.1 Cancro colorrectal

A patologia que serve de base ao presente estudo é o cancro colorrectal (CCR) - uma entidade nosológica do foro oncológico com incidência no cólon ou no reto, agrupada devido à relação anatómica das estruturas envolvidas bem como de outras características em comum.

Apesar de a prevenção e rastreio desta patologia ser acessível e difusamente aplicada, fazendo mesmo parte dos programas nacionais de rastreio obrigatórios, sendo este um factor cada vez mais apontado como método fundamental na redução não só da morbi-mortalidade mas também dos custos associados à doença, o CCR continua a apresentar-se como uma das entidades oncológicas com maior incidência no mundo ocidental.

Segundo Clerc et al. (2007) é o segundo tipo de cancro mais comum na Europa, quer entre homens quer entre mulheres. De acordo com Mar et al. (2017), esta patologia é a terceira mais frequente no País Basco entre o sexo masculino e a segunda entre o sexo feminino. Segundo Hovermann et al. (2011), o CCR é o quarto tipo de cancro mais diagnosticado nos Estados Unidos da América e a segunda principal causa de morte em doentes oncológicos. Outra característica, que parece ser aceite em toda a literatura, é que a

sua incidência aumenta com a idade, seguindo a tendência verificada para a maior parte dos tumores.

A patologia oncológica é diagnosticada através de métodos clínicos, associados a métodos imagiológicos e laboratoriais. Após o diagnóstico de cancro é importante a caracterização do tipo histológico do tumor (dependente das características anatomo-patológicas encontradas em produto de biópsia) assim como o estadiamento da doença, ou seja, a caracterização da extensão da doença no órgão de origem e a descrição da existência de doença em órgãos contíguos ou à distância. Existem quatro estadios referentes ao cancro colorrectal, identificados por numeração romana, de I a IV, em ordem de gravidade/extensão crescente. O avanço do estadio condiciona a redução da taxa de sobrevivência entre os pacientes oncológicos. A descrição pormenorizada destes estadios é irrelevante para o estudo em causa visto que não há informação no registo do Hospital da Luz sobre o estadio de cada paciente. Por este motivo não foi feita distinção entre o processo de tratamento consoante o estadio da patologia.

2.2 Clinical Pathway

Um dos conceitos centrais deste estudo é o método de standardização dos tratamentos aplicados à patologia selecionada e que assim torna a análise mais fácil e posteriormente comparável com análises do género em instituições que mantenham protocolos de tratamento semelhantes.

O termo Clinical pathway (CP) refere-se a um processo standardizado de tratamento para uma determinada patologia de acordo com as circunstâncias apresentadas pelo doente. O CP é desenvolvido por uma equipa multidisciplinar de peritos e com evidência clínica a suportar quer as diferentes tarefas nele incluído quer a sua respetiva ordem de aplicação.

Ao utilizar um CP, diferentes passos de um processo de tratamento de um doente podem ser previstos com maior precisão. Além disso, o CP pode ser usado para melhorar a logística e o processo de planeamento do departamento em questão. Subsequentemente, pode conduzir a uma melhor utilização do pessoal de saúde e de outros recursos, como por exemplo, dos equipamentos médicos existentes.

É de realçar que o processo descrito num qualquer CP não é vinculativo uma vez que nem sempre é o mais indicado para todos os doentes, sendo que o responsável clínico tem autonomia para aplicar os tratamentos que julgue necessários e mais adequados perante a situação individual de cada paciente, mesmo que essas ações diverjam do que é aconselhado no CP. No entanto, as indicações presentes num

CP existem porque, de acordo com estudos científicos sólidos, aumentam a qualidade do serviço prestado, a eficiência dos tratamentos e a utilização dos recursos necessários. Apresenta-se de seguida (Ilustração 2.1) um exemplo de parte de um CP.

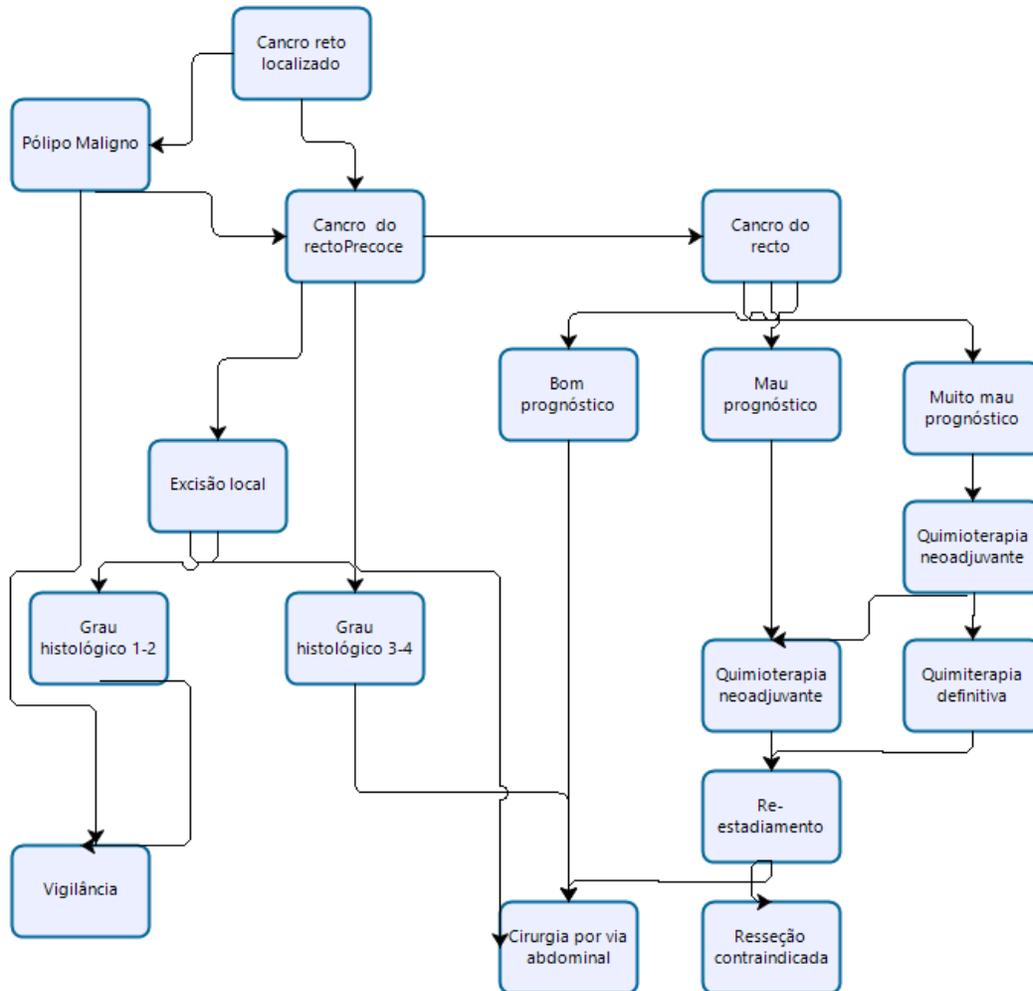


Ilustração 2.1 Diagrama do Clinical Pathway para o Cancro do Reto fornecido pelo Hospital da Luz Lisboa

2.3 Process Modeling

O *Business Process Management* (BPM) é uma corrente de gestão com propósito explicado de seguida e que engloba não só o *process modeling* mas também o *process mining*. A sua importância tem vindo a aumentar significativamente na área da saúde. Esta corrente de gestão recorre a modelos processuais para identificar, rever, validar, representar e comunicar a informação acerca de processos (Weske, 2007).

De entre a grande variedade de linguagens existentes para o efeito, o *Business Process Management Notation 2.0* (BPMN 2.0) é considerado como um padrão. Ainda assim pode ser, para domínios específicos, de difícil utilização. O sector da saúde é um ótimo exemplo devido à sua natureza multidisciplinar, em que os processos são habitualmente bastante complexos (Mans et al., 2008).

O BPMN é uma linguagem usada especificamente para a modelação de processos. Foi globalmente aceite nos círculos empresariais e institucionais muito graças ao facto de ter sido criada para ser usada e compreendida por qualquer utilizador, do administrador ao analista, do engenheiro ao perito de qualidade, sem qualquer necessidade de conhecimento técnico especializado. Esta linguagem fornece uma noção gráfica que expressa todos os aspetos do processo num único diagrama (Rojo et al., 2008). A notação utilizada cria uma ponte de ligação entre o *design* de processos e a implementação dos mesmos.

Como qualquer linguagem, o BPMN 2.0 utiliza símbolos específicos, os quais foram escolhidos para se diferenciarem entre si e serem formatos familiares para os modeladores (Ilustração 2.2). Existem, dentro da linguagem em causa, quatro categorias de elementos: *Pools and Lanes*, *Flow Objects*, *Connecting Objects* e *Artifacts* (White, 2004).

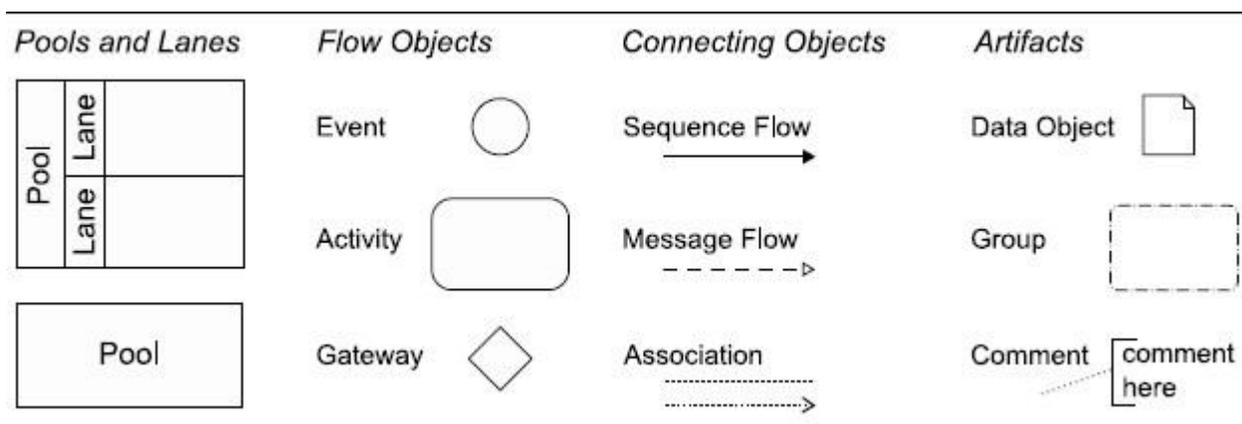


Ilustração 2.2 Elementos da linguagem BPMN <https://www.researchgate.net/>

As **Pools and Lanes** são usadas principalmente para ilustrar diferentes responsabilidades funcionais e utilizam dois tipos de elementos. As **pools** representam participantes no processo e são também usadas como um contentor gráfico para a divisão de atividades em outras **pools**, principalmente em contextos de *business to business* (B2B). As **lanes** são o segundo símbolo presente nesta categoria, sendo estas não mais do que uma sub-partição das **pools** e têm como objetivo principal a organização e categorização das

atividades. Apresenta-se em seguida um exemplo de um diagrama BPMN com uso de *lanes* para identificar dois diferentes atores.

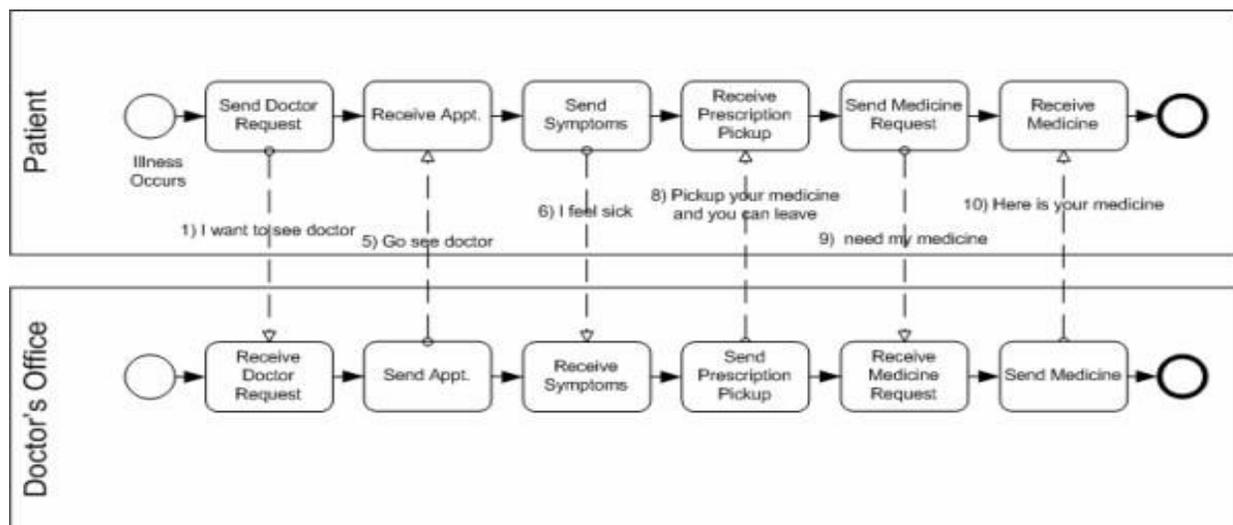


Ilustração 2.3 Exemplo de diagrama BPMN com lanes <https://www.stakeholdermap.com/bpm/doctor-to-patient-bpm.html>

A categoria **Flow Objects** inclui **eventos** que são representados por círculos. Estes podem marcar o início ou o final de um processo mas também existe a possibilidade de se encontrarem no interior do processo como evento intermédio. Os eventos impactam o processo porque estão habitualmente relacionados com uma causa ou um resultado. Adicionalmente figuram nesta categoria as **atividades**, representadas por retângulos com arestas de extremidades arredondadas. São usadas para descrever o que é efetivamente realizado no processo e podem ser do tipo atómicas ou subprocessos. A distinção é feita por um pequeno sinal “+” no lado inferior do retângulo que marca um subprocesso e que indica que aquela atividade compreende várias atividades. Finalmente, o último símbolo pertencente a esta categoria são as **gateways**, representadas por losangos (diamantes como são referidos na literatura). São usadas para controlar a convergência e a divergência do fluxo de sequência. Existem vários tipos de *gateway* consoante o tipo de decisão a tomar. A *gateway* pode ser exclusiva, obrigando à escolha de um dos vários caminhos opcionais, paralela, obrigando à execução simultânea dos caminhos que divergem dela e ainda inclusiva, permitindo a execução de qualquer caminho ou de mais do que um ao mesmo tempo.

A categoria **Connecting Objects** inclui três tipos de objetos que são usados para ligar os **Flow Objects**. Começando pela **sequence flow**, que é uma linha sólida terminando numa seta que indica a direção em

que o processo se vai efetuar, e serve para ligar diferentes atividades na ordem da sua execução. Dentro desta categoria existe também o **message flow**, uma linha tracejada com uma seta com o interior não preenchido e que representa a troca de mensagens entre participantes no processo. Finalmente, **association** é o último integrante desta categoria, sendo representado por uma linha pontilhada com uma seta no final. Este objeto é usado para associar dados, texto ou outro *Artifacts* a um qualquer *Flow Object*.

A categoria **Artifacts** é usada para adicionar algum contexto dada a situação dentro do processo modelado. Os símbolos presentes nesta categoria são **data objects**, representados por uma folha de papel com o canto superior direito dobrado e servem para indicar os dados necessários ou produzidos por uma atividade, encontrando-se ligados às atividades através de *association*. O segundo tipo de símbolo é designado **group**, sendo usado um retângulo com as extremidades das arestas arredondadas e linha tracejada para agrupar atividades, numa ótica de análise ou documentação, não produzindo qualquer efeito no fluxo de sequência. Por último existe ainda **annotation** que, como o nome indica, serve para deixar alguma nota para o leitor de um mapa de processos que não tenha ficado esclarecido com o uso dos outros símbolos (Muller et al. 2011). A ilustração 2.4 apresenta um segundo modelo em BPMN para um processo de forma a dar ao leitor uma noção visual do que é pretendido alcançar com esta linguagem. Nos anexos 2 e 3 são apresentados diagramas BPMN para o processo de tratamento do CCR.

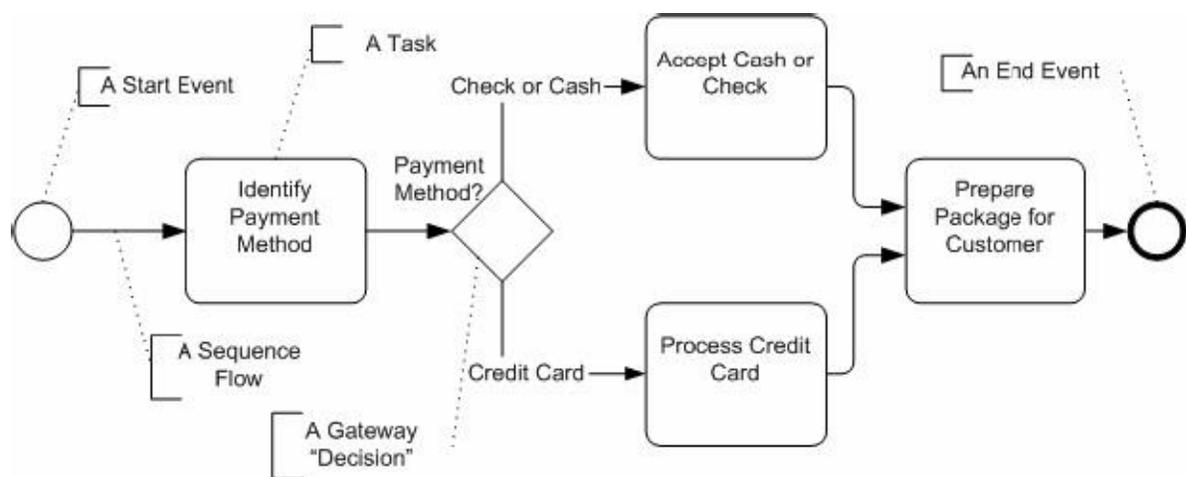


Ilustração 2.4 Exemplo diagrama BPMN com annotations <https://www.stakeholdermap.com/bpm/payment-process-bpm-large.png>

2.4 Event Log

O **Event Log** é o registo de todas as atividades de um processo e é essencial para a execução de *process mining*, sendo neste que se encontram os dados que serão processados. Assume-se a existência de um *event log* em que cada evento se refere a um caso em particular, a uma atividade e a um determinado momento no tempo. Pode ser também visto como uma coleção de casos e cada caso pode ser então tido como uma sequência de eventos (também designada por *trace*).

A informação dos *event logs* pode provir de diferentes origens, seja uma base de dados, um ficheiro em formato CSV (comma separated values) ou *spreadsheet* (originário em uma base de dados), um *transaction log*, um sistema de *Enterprise Resource Planning* ((ERP) por ex., SAP), um *message log* (*event log* para mensagens de onde pode ser extraída também valiosa informação sobre o processo em estudo) ou um API (*Application Programming Interface*) que recolhe informação de *websites* ou de redes sociais.

No estudo realizado, os *event logs* foram disponibilizados num ficheiro CSV, sendo necessário a consideração de alguns requisitos mínimos para o seu processamento. Em primeiro lugar, cada linha deve estar unicamente associada a um evento. Adicionalmente, devem existir pelo menos três colunas, que apresentem o identificador do caso (*case ID*), o nome da atividade e o *timestamp* (*startTime*). Há igualmente a possibilidade de existência de colunas adicionais com informação suplementar sobre os eventos em causa, como por exemplo, custos, idades dos utentes, recursos ou tipos de transação (Van der Aalst et al., 2012). A tabela 1 apresenta um extrato de um ficheiro CSV disponibilizado pelo Hospital da Luz em que se podem ver vários *event logs*.

Tabela 1 Exemplo de event log fornecido pelo Hospital da Luz

	A	B	C	D	E	F
1	case	event	startTime	Especialidade	Staff	Valor
2	100169716	MCDTs Imagiologia	2017/04/01 00:00:00.000	Imagiologia	42659	132,25
3	100169716	MCDTs Imagiologia	2017/04/01 00:00:00.000	Imagiologia	34923	11,5
4	100169716	MCDTs	2017/04/04 00:00:00.000	Patologia Clínica	96125	147,27
5	100169716	MCDTs Especialidades	2017/10/23 00:00:00.000	Anatomia Patológica	92260	46
6	100169716	MCDTs Especialidades	2017/10/23 00:00:00.000	Anestesiologia	91576	230
7	100169716	MCDTs Especialidades - Gastreenterologia	2017/10/23 00:00:00.000	Gastreenterologia	92260	517,79
8	100169716	Consulta	2017/11/15 00:00:00.000	Gastreenterologia	92260	40,25
9	100169716	MCDTs Imagiologia	2017/11/18 00:00:00.000	Imagiologia	91374	345
10	100169716	MCDTs	2017/11/20 00:00:00.000	Patologia Clínica	96125	66,7
11	100169716	Consulta	2017/11/27 00:00:00.000	Cirurgia Geral	91007	40,25
12	100169716	Consulta	2017/11/29 00:00:00.000	Oncologia Médica	30799	0
13	100169716	Consulta	2018/01/08 00:00:00.000	Cirurgia Geral	91007	40,25

2.5 Process Mining

O conceito de *process mining* foi desenvolvido para extrair uma descrição estruturada de um processo a partir de um conjunto de execuções reais (Aalst et al., 2003). Alternativamente, Fei et al. (2010) definem este conceito como um conjunto de técnicas, ferramentas e métodos para descobrir, monitorizar e melhorar processos reais através da extração de conhecimento dos *event logs*. No contexto do presente estudo, deve assim ser construído um modelo (em BPMN) apropriado com o propósito de avaliar e melhorar a eficiência dos serviços envolvidos no processo de tratamento de CCR.

As técnicas de *process mining* pressupõem que todos os processos possuem uma identificação única e que as ocorrências de tarefas/atividades estão registadas num *event log*. Existem três técnicas de extração de conhecimento (Fei et al., 2010): descoberta, análise de conformidade e melhoria, que se apresentam de seguida.

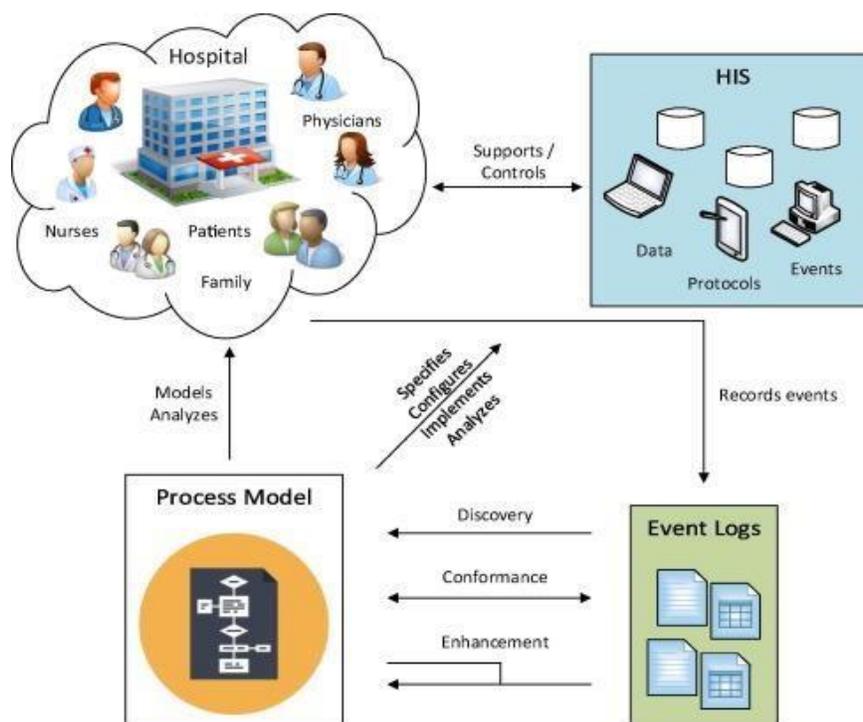


Ilustração 2.5 Diagrama explicativo de process mining em *Process Mining in Healthcare* (Rojas, 2016)

Descoberta: é a técnica de *process mining* mais utilizada. Usa o *event log* para produzir um modelo sem qualquer informação *a priori*. O output desta técnica é um modelo do processo, habitualmente em formato *Petri Net* (representação matemática para sistemas distribuídos discretos), BPMN ou UML

activity diagram (gráfico de fluxo usado para controlo).

Análise de conformidade: esta técnica é usada para comparar um modelo de um processo com o *event log* desse mesmo processo. Pode ser assim usada para averiguar se a realidade registada no *event log* está em concordância com o modelo teórico e vice-versa. Finda a análise de conformidade, o *output* contém um diagnóstico informativo sobre diferenças e semelhanças entre o processo e o *event log*.

Melhoria: com esta técnica idealiza-se melhorar o processo atual utilizando alguma informação registada no *event log*. O *output* desta técnica é um modelo melhorado.

Para além das atividades supracitadas, as técnicas de *process mining* podem possuir várias perspetivas. A **perspetiva de controlo de fluxo** (utilizada neste estudo) foca-se na ordem das atividades, com o objetivo de criar uma boa caracterização de todos os caminhos possíveis. No que toca à **perspetiva organizacional**, o foco está nas informações sobre os recursos. Por exemplo, permite identificar os atores que estão envolvidos e como estes se relacionam. Existe também uma **perspetiva centrada nas propriedades dos casos registados** que está ligada à extração de informações descritivas acerca dos casos individuais. Por último, existe ainda a **perspetiva temporal** que se concentra no *timing* e na frequência dos eventos. Pode avaliar-se o processo em termos de gargalos, níveis de serviço, utilização de recursos e prever o tempo de processamento restante para os casos atuais.

Apesar da complexidade dos processos de tratamento, o *process mining* tem vindo a ganhar importância nos últimos anos no sector da saúde. É um método de avaliação da correta aplicação prática das indicações presentes num CP. O domínio desta aplicação apresenta desafios e limitações, maioritariamente relacionados com (1) o facto de diferentes tipos de pacientes apresentarem diferentes características, (2) o estado dos utentes evoluir dinamicamente e, por último, (3) devido a diferentes hospitais possuírem diferentes restrições em termos de recursos.

Muitos algoritmos utilizados em *process mining* não têm capacidade de lidar com o ruído (dados incorretos ou em falta num *event log*) que costuma estar presente em processos reais, ficando a sua utilização por conseguinte limitada. Para além disso, a maioria dos métodos disponíveis não são customizados para aplicação médica e não têm em conta a necessidade de contextualização. Existem ainda algumas limitações comuns a muitos dos algoritmos mais aplicados nesta tipologia de avaliação, nomeadamente: (1) a extração de padrões dos processos sem contextualização, ou seja, o mesmo

procedimento pode ter efeitos diferentes em utentes distintos dependendo do contexto; (2) a mineração de processos que não correspondem a nenhuma sequência de eventos pertencente ao *event log* usado como input, neste caso os modelos produzidos pelo algoritmo são usados para avaliação de qualidade e para análise de conformidade, e os resultados produzidos podem ser duvidosos se o input não for confiável; (3) não relacionam explicitamente os padrões minados com o *event log*; esta relação pode ser importante não só para caracterizar o contexto mas também para providenciar aos profissionais de saúde evidências acerca do *output* obtido (Canensi et al., 2016). Existem também outros desafios no que toca aos dados a utilizar e a certas propriedades menos benéficas que estes podem apresentar, tais como, estarem distribuídos por várias fontes, o que se torna particularmente complicado quando essas fontes distintas usam diferentes identificadores. Adicionalmente, os dados podem (1) estar centrados no objeto ao invés de no processo; (2) estar incompletos; (3) incluir *outliers* que conferem comportamento excecional ao processo e a possível verificação de diferentes granularidades em eventos de um mesmo *event log*. Uma outra consideração a ter quanto aos dados a utilizar refere-se ao número de registos do *event log*, isto porque, em situações extremas, *event logs* com dimensões demasiado grandes poderão ser difíceis de processar enquanto dimensões demasiado pequenas podem não conter informação suficiente. Deve-se ainda ter em conta o número médio de eventos por caso, a semelhança entre casos e o número de eventos ou caminhos únicos.

Dado que esta disciplina é recente e ainda se encontram algumas limitações e desafios, existem certos princípios que podem guiar o utilizador a melhores resultados. Ter em atenção a qualidade dos dados é essencial, uma vez que os resultados vão depender da qualidade do *input*. Assim, é suposto que os dados sejam confiáveis, completos e com semântica bem definida. Infelizmente os *event logs* são frequentemente um subproduto usado meramente para depuração e perfilhamento. É então de extrema importância que a extração do *event log* seja feita com as perguntas de investigação em mente e que a escolha da representação seja corretamente feita de forma a possibilitar a obtenção de resultados congruentes resultantes da aplicação de modelos apropriados (Van der Aalst et al. 2012).

2.6 Análise de Conformidade

A análise de conformidade determina a que nível o comportamento do modelo é refletido no *event log*. Um doente que siga perfeitamente o CP atinge o nível de conformidade 1, já um utente cujo tratamento

se desvie do caminho previsto no CP terá um valor inferior de conformidade. Evidentemente, quanto mais um processo de tratamento se desviar do processo estandardizado, menor será a conformidade registada sendo que na situação de desfasamento total entre o CP e o *event log* a conformidade terá valor 0. A análise de conformidade baseia-se em quatro critérios de qualidade que se concentram nas quatro seguintes questões (Van de Steeg, 2015):

Fitness: “O *event log* reflete o comportamento apresentado no modelo?”.

Precisão: “Existem, no *event log*, comportamentos alheios ao observado no modelo?”.

Generalização: “Existe generalização do modelo quanto ao observado no *event log*?”.

Simplicidade: “Quão simples é o modelo?”.

Para além de verificar a concordância entre o *event log* e o CP, a análise de conformidade permite também retirar outras informações. Por exemplo, o estudo efetuado no Hospital Isala em Zwolle (Holanda), pretendeu averiguar se o nível de conformidade tem influência na ocorrência de incidentes e nos custos. No caso da não conformidade resultar em custos mais elevados e no aumento do número de incidentes, é necessário intervir para assegurar a aderência às indicações do CP. Por outro lado, se a não conformidade levar a custos mais baixos e/ou a menos incidentes, poderá ser investigado quais os desvios que são realmente benéficos (Van de Steeg, 2015).

2.7 ProM

ProM (abreviatura para *Process Mining framework*) é um software *open source* que contém algoritmos de *process mining*. O software fornece uma plataforma para utilizadores e criadores dos algoritmos de *process mining* que é simples de usar e facilmente extensível. A missão aliada à criação deste *software* é torná-la a plataforma padrão no mundo académico, estabelecendo uma comunidade ativa e reconhecida de utilizadores e contribuintes e, além disso, consciencializar a generalidade das pessoas envolvidas em gestão de processos para as valências da tecnologia de *process mining*, através da promoção da sua aplicação e da aceitação industrial. A visão dos criadores é avançar o estado de arte da tecnologia, desenvolvendo métodos funcionais, criando uma comunidade aberta e providenciando uma plataforma extensível que suporta a tecnologia aqui mencionada.

Esta plataforma foi desenvolvida pela Universidade de Eindhoven (Holanda), atualmente na versão 6.8, e

apresenta-se como uma das mais usadas na área da saúde. Conta com centenas de *plug-ins* com diferentes funcionalidades, desde *plug-ins* de descoberta, a *plug-ins* de melhoria e passando por *plug-ins* de análise de conformidade. A navegação no software é bastante intuitiva, estando qualquer das suas funcionalidades disponibilizada de forma rápida e bastante acessível. Foi a plataforma escolhida quer pelas qualidades já referidas, quer pelo reconhecimento da sua valia observado na presença constante em literatura relacionada com o tema atesta.

2.8 Técnicas alternativas de análise de conformidade

Este ponto visa apresentar técnicas de análise de conformidade retiradas da literatura que se consideram alternativas à utilizada neste estudo. Apresentam-se características e uma breve explicação do funcionamento de cada uma.

Weidlich et al. (2011) apresentam um conjunto de medidas de conformidade, não especificamente focadas no sector da saúde, mas aplicáveis a qualquer processo desde que modelado em linguagem BPMN. Funcionam com recurso a um conjunto de restrições comportamentais que um modelo processual impõe a pares de atividades ao invés de tentar replicar os casos através de uma noção rigorosa de equivalência. Cada par de atividades está, no mínimo, associado a um dos seguintes tipos de relação: relação de ordem estrita (uma atividade sucede a outra), relação de intercalação (uma atividade pode acontecer antes ou depois de outra), atividades exclusivas (quando duas atividades não podem ocorrer numa mesma sequência) e relação causal (a realização de uma atividade força a ocorrência de outra). Na ótica dos autores enquanto alguns casos são realizações válidas, outros evidenciam apenas parte das execuções recomendadas, focando-se em casos onde o modelo processual é normativo.

Milchak et al. (2006) definem um conjunto compreensivo e válido de critérios para um processo que reflita o JNC 7 (práticas recomendadas para tratamento de hipertensão) e desenvolvem um método de pontuação que permite a caracterização quantitativa da aderência dos profissionais de saúde ao JNC 7. Para este efeito, os investigadores elaboraram, com o auxílio de especialistas na área da patologia, uma lista de 22 critérios agrupados em 4 domínios gerais do tratamento da hipertensão: diagnóstico, terapia medicamentosa, intensidade de acompanhamento e monitorização laboratorial. Para além disso foram associados diferentes pesos aos critérios refletindo a importância dos mesmos. Foram atribuídos os valores 1 e 2 aos critérios com base na opinião dos especialistas. O algoritmo associado a este método

efetua, com a aplicação de uma percentagem acima da qual se considera uma aderência maioritária a um critério, a contabilização pontual de um *event log*, refletindo assim o nível de conformidade do mesmo. A seguinte metodologia de análise de conformidade de *guidelines* clínicas em relação a um processo real é denominada “*de jure model*”. Foi proposta por Rovani et al. (2015) e utiliza a linguagem declarativa para modelação. Através de um conjunto de técnicas de mineração processual (*Declare Analyser* e *Declare Checker*) suportadas pelo software ProM Tools, os autores identificam desvios e propõem ações a tomar para adaptar e melhorar o CP. O método apresentado faz a análise de conformidade e a codificação de indicações clínicas automaticamente. Depois, repara o “*de jure model*” para ser consistente com a conduta real produzindo o “*de facto model*” baseado em validação cruzada. O algoritmo separa o *event log* em dois, uma parte faz as reparações no “*de jure model*” e a outra serve para analisar a conformidade. Se o nível de conformidade for satisfatório, o modelo reparado é mantido e considerado o “*de facto model*”, caso contrário o processo é repetido até que um valor de conformidade satisfatório seja obtido.

Como último exemplo de metodologias de análise de conformidade surge o estudo de Huang et al. (2014) que permite identificar em tempo real violações do procedimento padrão e elabora a respetiva análise de conformidade. Este trabalho apresenta um sistema de análise de conformidade baseado em regras de semântica, mais concretamente, construindo uma ontologia para o CP. As restrições de tratamento são modeladas em *Semantic Web Rule Language* (SWRL) para especificar os processos idealizados de tratamento e a quantificação da estrutura temporal do CP. A metodologia proposta fornece uma análise de conformidade em paralelo com o tratamento do utente em vez de uma análise *a posteriori*. Os resultados demonstram que esta metodologia se aproxima de uma análise de conformidade em tempo real e permite aos intervenientes a deteção de desvios e a tomada de decisão informada nas situações de não conformidade ainda durante a execução do processo de tratamento.

CAPÍTULO 3: METODOLOGIA

Neste capítulo apresenta-se a metodologia adotada durante o estágio realizado com o intuito de efetuar a análise de conformidade no processo de tratamento do cancro colorretal no Hospital da Luz de Lisboa, e o estudo do impacto do grau de conformidade na variação de custos associada.

Este capítulo divide-se em 5 tópicos e cada um reflete uma parte prática deste estágio começando por uma breve descrição da metodologia empregue no ponto 3.1, a descrição do processo de recolha de dados no ponto 3.2, a abordagem do processamento dos dados no ponto 3.3, abordando no ponto 3.4 a modelação do processo com ligação ao último ponto (3.5) que se dedica ao *process mining*.

3.1 Técnica de pesquisa

Para suportar teoricamente o presente estudo foi realizada uma pesquisa bibliográfica. Para isso, com as perguntas de investigação em mente iniciou-se uma pesquisa no Google académico e que teve por base a introdução das seguintes palavras-chave: “Colorectal Cancer”, “Clinical Pathway”, “Process Mining”, “Process Modelling”, “Healthcare processes”, “BPMN (Business Process Management Notation)”, “ProM Tools”, “Compliance checking” e algumas junções entre as selecionadas. Foram pesquisados termos em inglês devido ao ramo de *process mining*, em particular na área da saúde, ser ainda bastante recente e pela maior parte dos artigos desta área estar escrito em inglês.

3.2 Dados

Na versão inicial, os dados anonimizados continham o registo financeiro dos gastos associados ao tratamento de 116 utentes com diagnóstico da patologia em estudo (CCR) que foram tratados no hospital em 2017 e 2018. Para além do número de identificação (ID case) de cada paciente, os dados continham a data de cada gasto (em atividades, em materiais ou outras necessidades para a execução de um procedimento), a especialidade associada, o agrupador que é não mais do que o conjunto dos referidos gastos que podem ou não ser relativos a procedimentos, o *staff ID* (isto é, a identificação do executante de cada procedimento), a quantidade e o valor. Em 3.3 descrevem-se mais aprofundadamente os dados e especialmente o processamento necessário para se poder realizar um *process mining*.

3.3 Processamento dos dados

O processamento dos dados é essencial para a sua utilização em *process mining*. Como foi referido em pontos anteriores, existem certas condicionantes para o uso de dados provenientes de sistemas de informação que não são exclusivamente dedicados ao processo em estudo. Este facto está presente neste estudo pois o sistema de informação do hospital foi construído para responder a exigências de uma perspetiva financeira. Para além de incluir os procedimentos médicos (que são as atividades a investigar), também contém materiais e fármacos utilizados nessas mesmas intervenções e, conseqüentemente, surge a necessidade de realizar uma primeira triagem da informação. Nesta primeira fase de processamento, é necessário retirar as linhas de excel que se referem a recursos usados durante procedimentos da base de dados, afetando os custos das atividades a que esses mesmos recursos estavam associados. São retirados porque não representam atividades e por isso não devem estar no registo usado para a mineração que vai ter como perspetiva o controlo de fluxo.

O ficheiro excel com o registo associado ao processo, fornecido pelos serviços de informação do Hospital da Luz de Lisboa, continha 20836 linhas e foram limpas 9733 referentes a “bloco”, “materiais” e “fármacos” cujos custos foram, como supracitado, afetos às atividades com que estavam relacionados. Ficam assim 11103 linhas referentes a procedimentos a que os utentes presentes na base de dados foram sujeitos. Usaram-se os filtros do excel para selecionar as linhas que necessitavam de ser retiradas e a consola (VBA) para retirar as linhas vazias do *event log*.

A segunda fase de pré-processamento dos dados relacionou-se com uma standardização da nomenclatura das atividades. A elevada pormenorização iria dificultar a análise pelo que houve a necessidade de uniformizar certas categorias de atividades. Tal verifica-se, principalmente no que dizia respeito às intervenções cirúrgicas, que estavam catalogadas com o nome da intervenção, resultando em mais de duas dezenas de intervenções diferentes, o que, para a análise em causa, não seria necessário. Inicialmente eram referidas 133 diferentes atividades que foram agregadas em apenas 12 atividades diferentes. Posteriormente, eliminaram-se também atividades com a mesma designação e com igual *timestamp*. Por último, foram identificadas, através do ID de cada utente, todas as atividades que estão diretamente relacionadas com o tratamento do cancro do colorretal. No final desta etapa de

processamento subsistem 5920 linhas (28%).

A última fase de pré-processamento foi realizada com o software ProM Tools 6.8, com recurso ao algoritmo *Filter Log Events*, que permitiu a seleção de casos que tivessem como eventos iniciais e finais “Consulta”, “MCDTs”(métodos complementares de diagnóstico e tratamento), “MCDTs Imagiologia” e “MCDTs Gastroenterologia”. Desta forma, garantiu-se que as atividades iniciais e finais eram congruentes com o que se verifica no *clinical pathway* e que os casos em que os utentes realizaram todo ou grande parte do tratamento referente à patologia em estudo no Hospital da Luz foram selecionados. Dos 116 casos iniciais ficaram 59, 56 depois de retirados os três *outliers* (estes foram selecionados por serem casos com excessivo número de atividades (superior a 250) tendo em conta o resto da amostra). Das 5920 atividades registadas no *event log* importado para o ProM restaram, depois deste último processamento, 2099.

3.4 Modelação processual do Clinical Pathway

Neste ponto discute-se a modelação do CP de forma a poder realizar-se a análise de conformidade entre o *event log* e o processo standardizado. Este foi, depois da revisão da literatura para ter alguma familiaridade com o sector e o próprio tema, o primeiro passo durante o estágio. O CP, em aplicação atual, foi fornecido pelo Hospital e na forma em que é facultado aos profissionais de saúde. O facto de o CP fornecido pelo Hospital da Luz ser elaborado para profissionais de saúde criou a necessidade de aprofundar o conhecimento através de pesquisa para dessa forma retirar o essencial da informação nele presente, tornando-o, ao mesmo tempo, perceptível para a generalidade dos leitores. Apesar de estar já parcialmente mapeado, o CP distinguia tratamentos para cancros em estadios diferentes (informação que não foi incluída no registo facultado). Para além das distinções em termos de estadiamento, o *pathway* estava descrito numa linguagem pouco perceptível aos utilizadores comuns, existindo assim a necessidade de generalizar os termos usados de forma a que intervenientes menos familiarizados pudessem retirar o essencial da análise.

Numa primeira fase a ideia foi transpor o que estava originalmente mapeado no CP para linguagem BPMN usando o software BIZAGI. O resultado foi o mapa apresentado no anexo 2 que tem a particularidade de

ter quatro *lanes* correspondendo às fases de tratamento descritas no *pathway* (Diagnóstico, Estadiamento, Tratamento e Vigilância). Este modelo é excessivamente complexo e contém demasiada informação. Portanto, numa segunda versão, o objetivo foi simplificar o modelo do processo, principalmente no que toca à parte do tratamento que estava, como dito anteriormente, especificada por estadio da patologia e, de acordo com o observado nos dados facultados, não seria possível para a análise. Para isso criaram-se fluxos que pudessem ser comuns a utentes com diferentes estadios, ou seja, criaram-se caminhos mais generalizados (generalização validada pelo Hospital), que incluem as atividades de tratamento comuns aos doentes. Esta fase realizou-se através da restrição da granularidade da informação, ou seja, dentro de alguns tipos de atividade, como por exemplo as terapias radiológicas e químicas, que seriam diferentes para utentes com patologias em estadios distintos, foram generalizadas. O novo mapeamento inclui apenas quimio e radioterapia que são unicamente distinguidos pela posição assumida em relação a procedimentos cirúrgicos (neo-adjuvantes no caso de precederem a uma cirurgia e adjuvantes na situação oposta).

Na terceira e última fase da modelação foi agregada toda a informação que estava presente na versão anterior numa só *lane*, já que, como referido em pontos anteriores, a informação disponibilizada e mesmo a dinâmica e complexidade deste tipo de processos (para além da singularidade dos pacientes) incutem uma dificuldade extra em identificar claramente a fase de tratamento em que o utente se encontra. Por outras palavras, não há um claro benefício em manter essa distinção e adicionalmente verifica-se um ganho de simplicidade no modelo. Este é um dos pontos centrais quando se trata do uso de *Process Modeling*. O presente modelo pode ser visualizado no anexo 3, e será a referência *a posteriori* do que é expectável em relação ao processo de tratamento, isto é, será o termo de comparação para o *event log* na posterior análise de conformidade.

3.5 Mineração do *event log*

Este ponto apresenta a técnica utilizada no presente estudo para chegar a valores de conformidade entre o que foi registado pelo Hospital e o CP atualmente em prática no mesmo. A utilização do algoritmo explicado de seguida foi inspirada num trabalho previamente desenvolvido no Hospital da Luz de Lisboa também sobre conformidade em processos de saúde, mais concretamente, sobre a mesma patologia

(CCR) (Barros, 2018). A diferença do presente estudo está na análise mais económica, na perspetiva de verificar o impacto do nível de conformidade nos custos e não focando só a análise de conformidade em si. Os resultados dos estudos quanto ao nível de conformidade foram distintos porque:

(1) usaram dados de períodos temporais diferentes; (2) usaram modelações distintas; (3) o processamento de dados foi feito em perspetivas diferentes em cada estudo. O artigo supracitado foi de qualquer forma uma referência na estruturação do estudo prático e também fonte de referências bibliográficas importantes sobre o tema.

Algoritmo aplicado na análise de conformidade

À medida que os processos se tornam mais complexos e dinâmicos a importância de conseguir elaborar bons modelos processuais aumenta. Muitas organizações procuram também encontrar um equilíbrio entre flexibilidade (os profissionais podem desviar-se do procedimento padrão) e controlo (os desvios têm de ser monitorizados e corrigidos se necessário). Daí o crescente interesse em análises de conformidade. Uma análise de conformidade mede a compatibilidade do *event log* com o modelo estabelecido. Existem vários critérios para avaliar conformidade sendo que o mais importante é sem dúvida o de *fitness* e o algoritmo aplicado foca-se única e exclusivamente neste critério para avaliar a conformidade.

Dado um modelo e uma sequência de um *event log* (caso), a *trace fitness* é alta caso a mesma sequência, ou uma muito semelhante, seja possível de acordo com o modelo. Os registos podem não conter todas as atividades executadas, mas ainda assim a sua identificação é importante porque alteram o desempenho da execução e podem conduzir a resultados não fiáveis.

Dado os inputs necessários à análise de conformidade (*event log* e modelo), desvios na dimensão de *fitness* manifestam-se como atividades *skipped* e *inserted*. *Skipped* são atividades que deveriam ser executadas segundo o modelo, mas que não ocorrem no registo, alternativamente *inserted* são atividades que ocorrem na realidade, mas que não estavam previstas no modelo.

Formas clássicas de medição de *fitness* penalizam a conformidade pela existência de qualquer um dos tipos de desvios acima mencionados. Adicionalmente não consideram severidades diferentes das atividades *skipped* ou *inserted*, o que resulta num valor de *fitness* que não corresponde ao nível de conformidade. O método desenvolvido por Adriansyah et al. (2011) propõe uma “*cost-based replay*

technique”, ou seja, uma técnica de replicação baseada em custos atribuídos às atividades (*skipped* e *inserted*) de modo a, com o contributo de uma função de custo que vai ser explicada mais adiante, calcular a *fitness* dos diferentes casos e do *event log* na sua totalidade. Para isso será necessário, antes de mais, definir alguns conceitos básicos.

O conjunto das atividades *skipped* são identificadas como A_s em A , em que A representa o conjunto das atividades do processo. As atividades *inserted*, por seu lado, manifestadas como eventos, são identificadas pelo conjunto $E_i \subset E_c$, em que E_c designa a sequência de eventos no caso c (instância), considerando $c \in C$ o identificador do caso (utente) dentro de um conjunto finito de casos. A metodologia permite a atribuição individual de custos às atividades *skipped* e *inserted* considerando a influência que elas possam ter no CP. k^s e k^i representam as funções de custo de “*skipping*” e “*inserting*” das atividades, respetivamente. A função apresentada de seguida define *fitness* (f) como um menos o rácio entre o custo total de ter atividades *skipped/inserted* e o custo total quando se considera que todos os eventos são atividades *inserted*.

$$f = 1 - \frac{\sum_{a \in A_s} k^s(a) + \sum_{e \in E_i} k^i(\alpha(e))}{\sum_{e \in E_c} k^i(\alpha(e))}$$

na qual $\alpha: E \rightarrow A$ representa a função que relaciona cada evento (e) do caso com a atividade do processo.

De um modo geral espera-se que um maior número de atividades *inserted* ou *skipped* faça decrescer o valor de *fitness*. No pior caso possível temos a situação de todos os eventos serem *inserted* que é usado para normalizar a métrica de *fitness*. Pela necessidade de saber *a posteriori* que atividades são *skipped* ou *inserted*, quando se identificam os desvios ao CP, o interesse prende-se em descobrir as atividades (*skipped* ou *inserted*) que têm menor custo e têm assim um menor impacto no valor de *fitness*.

De modo a replicar um grupo de eventos num modelo, as instâncias (*traces*) são iterativamente construídas usando prefixos dos casos. Para cada par que consiste num prefixo e em uma instância construída faz-se corresponder os eventos às instâncias de transição do processo que representam a mesma atividade. Cada vez que essa transição é observada admite-se a correspondência a um evento no processo. Quando consideramos eventos num prefixo para o qual não se encontra uma correspondência, significa que encontramos uma atividade *inserted*. Por seu lado consideram-se atividades *skipped* quando instâncias do processo não estão associadas a nenhum evento dos prefixos.

As funções de custo (k^s e k^i) são usadas para determinar as atividades das instâncias com maior tendência para causar os desvios. O objetivo passa, dessa forma, por construir instâncias que mostram o custo mais baixo dos desvios, de modo a que, unicamente em casos onde o custo de *skipping* seja superior a assumir uma atividade *inserted*, se considere uma atividade como *inserted*. Os custos associados às funções (k^s e k^i) serão os utilizados pelos algoritmos explicados a seguir para a criação de um grafo no qual se possam utilizar as suas propriedades para determinar que casos sofrem de menor ou maior impacto de atividades *inserted* ou *skipped*.

Na versão do ProM Tools adotada neste estudo (ProM Tools 6.8), quando se utiliza a metodologia de Adriansyah et al. (2011), que originalmente usava o algoritmo A* (algoritmo originalmente utilizado para a resolução do problema do caminho mais curto) como pesquisa, o software propõe dois algoritmos alternativos. Depois de experimentados, acabou por se selecionar o Dijkstra *based replayer*, ou seja, um replicador baseado no algoritmo de Dijkstra que, tal como o algoritmo A*, permite determinar o caminho mais curto e é aliás um caso particular do algoritmo A*.

O algoritmo A* tem uma função de custo definida por:

$$F(x) = g(x) + h(x)$$

Em que :

$g(x)$: função que define o custo real para atingir cada nodo x (tal como no algoritmo de Dijkstra).

$h(x)$: custo aproximado do nodo x para o nodo objetivo (no caso de localização é o nodo final). É uma função heurística que nunca deve sobrestimar o custo. Significa isto que o custo real de chegar de x ao nodo objetivo é sempre maior ou igual ao proposto nesta função.

O algoritmo A* só expande para um nodo que considere promissor. O seu único foco é chegar ao nodo objetivo. O algoritmo de Dijkstra é um caso particular do algoritmo A* quando $h(x) = 0$. Isto significa que não utiliza uma heurística e que a cada passo escolhe nodos com o menor custo, tende também, comparativamente a cobrir mais espaço no grafo (pesquisa menos localizada).

Dada esta característica, o algoritmo pode tornar-se mais útil em situações onde existem vários nodos objetivo concorrentes e em que não conhecemos o mais próximo. Caso fosse utilizado o algoritmo A* nesta situação seria necessário correr o algoritmo várias vezes (individualmente para cada novo

candidato). Isto justifica o uso do algoritmo Dijkstra no presente estudo. Na adaptação feita para este *replayer* os nodos são as atividades.

Relativamente a aplicação propriamente dita começa-se por abrir o ProM Tools 6.8, e seleccionar o botão *import* para importar o modelo em linguagem BPMN e o *event log* em formato CSV. É necessário transformar ambos os ficheiros em formatos compatíveis com o que é requerido pelo algoritmo de análise. Para isso usa-se o “*convert CSV to XES*” para usar o XES como formato para o *event log*. Adicionalmente, usa-se o “*convert BPMN diagram to Petri Net (control flow)*” para se obter uma *Petri Net* a partir do processo (BPMN) modelado em Bizagi. Os conversores encontram-se na lista de *plug-ins* disponíveis no ProM, são automaticamente sugeridos pelo software quando são detetados os formatos iniciais, mas podem também ser procurados na barra de procura no separador de ação. Faz-se o *upload* do modelo e do *registo* nos seus formatos mais recentes (XES e Petri Net) para o separador de ação e, mais uma vez, são sugeridos vários *plug-ins*, uns aparecem a cor verde querendo dizer que estão disponíveis e os que aparecem a amarelo indicam que falta algum elemento para a sua utilização. Seleciona-se dessa forma o *plug-in* aqui apresentado “*replay a Log on a Petri Net for conformance analysis*” de Adriansyah (2011) para a análise de conformidade. Ao iniciar o *plug-in* no software é indicado que se vai medir o *fitness* e pergunta se se penaliza ou não a conclusão imprópria. No corrente estudo considera-se que se penaliza e selecciona-se o primeiro algoritmo proposto para esse efeito, também referido neste ponto “*Dijkstra based replayer*”. Ao avançar é necessário especificar custos para as atividades. Optou-se por usar o custo unitário tanto para “*move on model*” (o que se verifica no modelo não corresponde ao *event log*) como para “*move on log*” (evento aparece no *event log* mas não tem correspondência no modelo). Passado este ponto executa-se o algoritmo. Os resultados obtidos são analisados no capítulo 4 de um modo geral e a respetiva interpretação destes no capítulo 5.

CAPÍTULO 4: RESULTADOS

Nesta secção apresenta-se a análise de resultados que foi desenvolvida durante o estudo. Foram estudados não só os custos, mas também níveis de conformidade e desenvolvidos alguns métodos para estudar relações entre variáveis distintas. Para isso criou-se a tabela presente em anexo 1. Esta tabela relaciona o caso particular de cada utente com os gastos associados a este para cada atividade, ficando

assim também registado, analogamente, o custo total por atividade. Até aqui os valores referidos são resultado de agregação de valores contidos no *event log* e somas simples de linhas/colunas (gastos por doente/gastos por atividade). De seguida foi calculado o peso de gastos por atividade no total dos gastos, obtendo os resultados apresentados na ilustração 4.1.

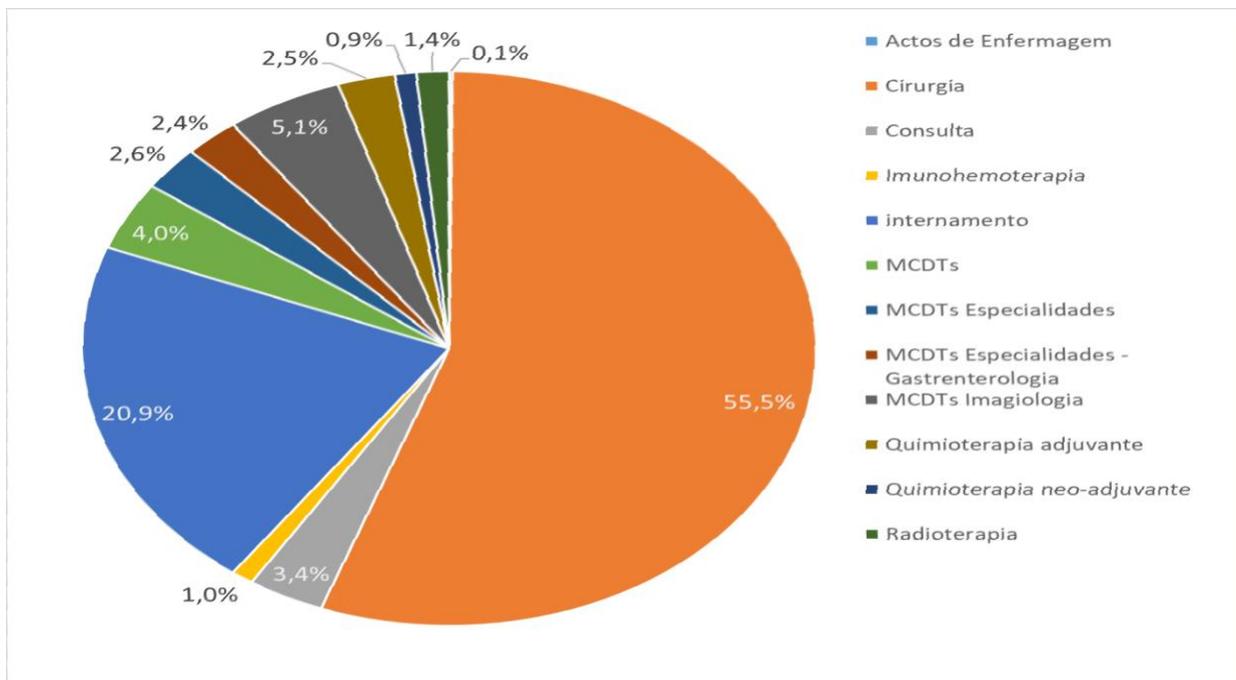


Ilustração 4.1 Gráfico circular com o peso em percentagem do custo de cada atividade em relação ao custo total

Observa-se que cirurgia (55,5%) e internamento (20,9%) representam mais de três quartos de todos os gastos, enquanto tratamentos químicos e radiológicos representam quase 5%. Os métodos complementares de diagnóstico e tratamento representam pouco mais de 14%. O preço cobrado em instituições de saúde privadas pode ser uma possível explicação para o baixo valor registado em terapias químicas e radiológicas e também em MCDTs. Esta hipótese reforça-se com o facto de as cirurgias representarem a grande parte dos gastos o que, poderá ter a ver com a vontade dos utentes em evitar as longas filas de espera que existem no sector público.

De seguida foi calculada uma média ponderada para os gastos por atividade e por doente. Ponderada porque foram contabilizados apenas os utentes que registaram gastos nas respetivas atividades e não em toda a amostra. Os valores finais podem ser consultados na tabela do anexo 1. Os valores de *trace fitness*

para cada caso (utente) foram também acrescentados à tabela supracitada. Na ilustração 4.2 apresenta-se o *box plot* que apresenta a variação dessa variável na amostra.

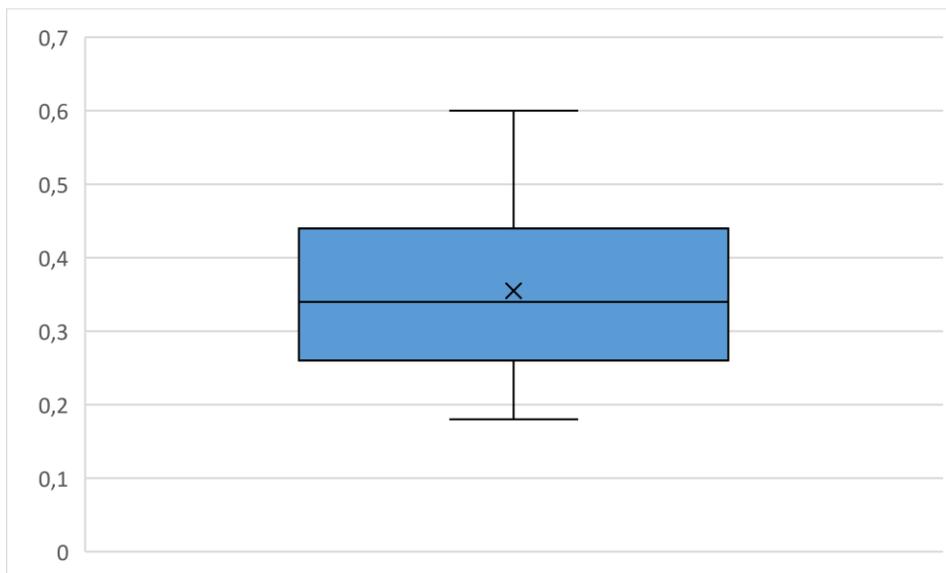


Ilustração 4.2 Box plot da trace fitness

Pode-se verificar que os valores de *trace fitness* se situam no intervalo [0,18; 0,60], com média de 0,3548, e mediana igual a 0,34. O intervalo que compreende os quartis Q_1 e Q_3 é [0,26; 0,44], 75% da amostra tem *trace fitness* igual ou inferior a 0,44 o que nos mostra, mais uma vez, que os resultados desta métrica estão longe de ser satisfatórios.

Posteriormente, elaborou-se um diagrama de dispersão para estudar a relação entre o *trace fitness* e o número de atividades de cada paciente, de forma a tentar perceber se existia ou não uma relação observável.

Verifica-se no diagrama abaixo que existe uma relação potencial entre o *trace fitness* e o número de atividades com a equação visível na ilustração e coeficiente de determinação (r^2) de 0,8141, o que expressa uma relação bastante forte. Pode-se visualizar e concluir que, nesta amostra, um maior número de atividades está associado a um nível de conformidade (*trace fitness*) menor.

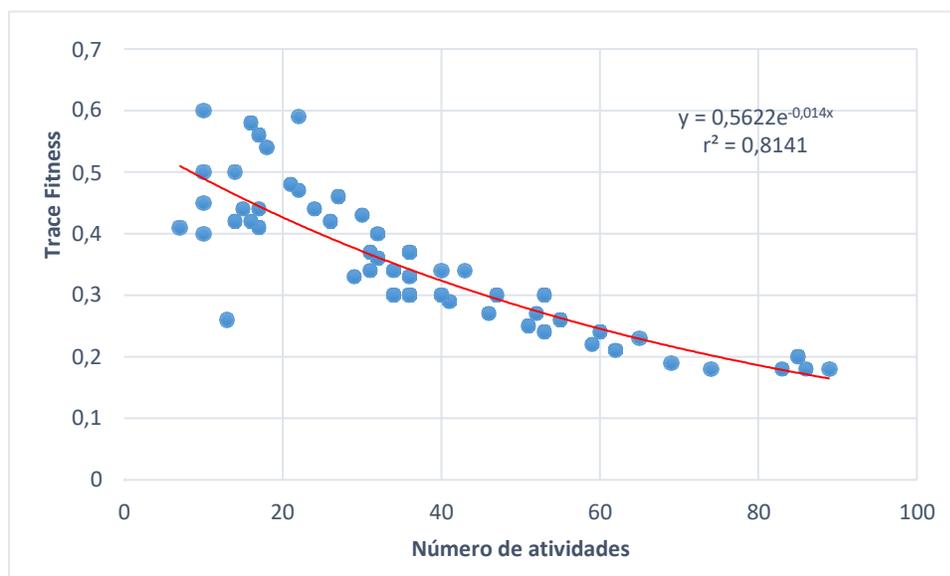


Ilustração 4.3 Diagrama de dispersão entre o trace fitness e o número de atividades

Da mesma forma verificou-se qual era a relação entre *trace fitness* e o custo total de cada paciente e entre *trace fitness* e o desvio absoluto em relação ao custo do processo standard (CP).

O desvio absoluto foi calculado através da subtração ao custo total por utente do valor do (CP). Este foi estimado a partir do cálculo da média do custo individual de cada atividade com recurso à informação contida no *event log*, somando o custo para cada um dos diferentes caminhos que são possíveis no modelo. Um dos caminhos possíveis não prevê a realização de ato cirúrgico e os outros dois distinguem-se pela realização de quimioterapia neo-adjuvante ou não. Acabou por se observar a existência de três custos diferentes, correspondentes às três alternativas de caminhos presentes no CP. Tendo em conta o caminho com que cada caso se relaciona melhor (observando as atividades a que o utente foi sujeito, principalmente nas atividades cirurgia e a existência ou não de tratamentos neo-adjuvantes) calculou-se uma média ponderada com valor 8293,16€. Este valor passou a ser a base de comparação de cada caso com o CP. De seguida apresentam-se, as duas ilustrações referentes aos dois diagramas de dispersão referidos neste parágrafo.

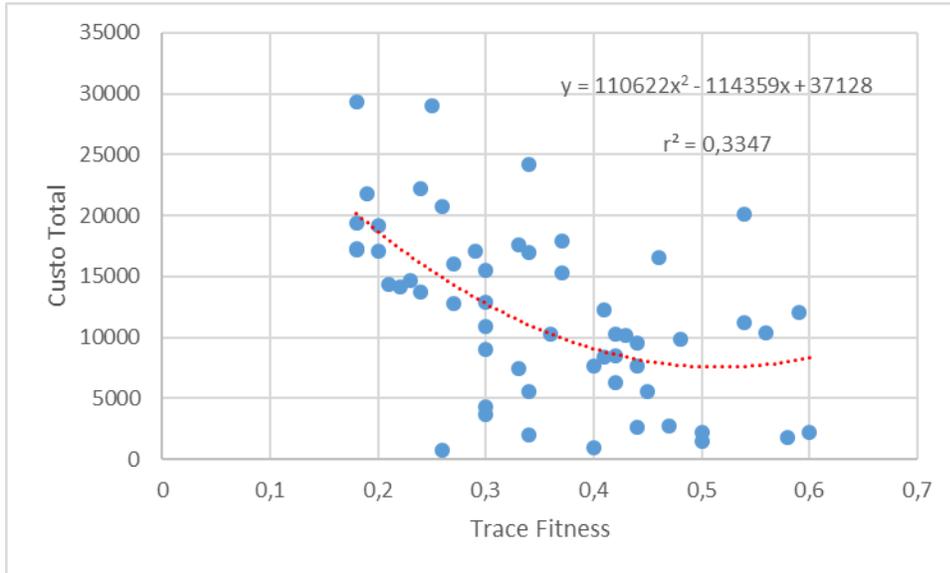


Ilustração 4.4 Diagrama de dispersão entre o trace fitness e o custo total

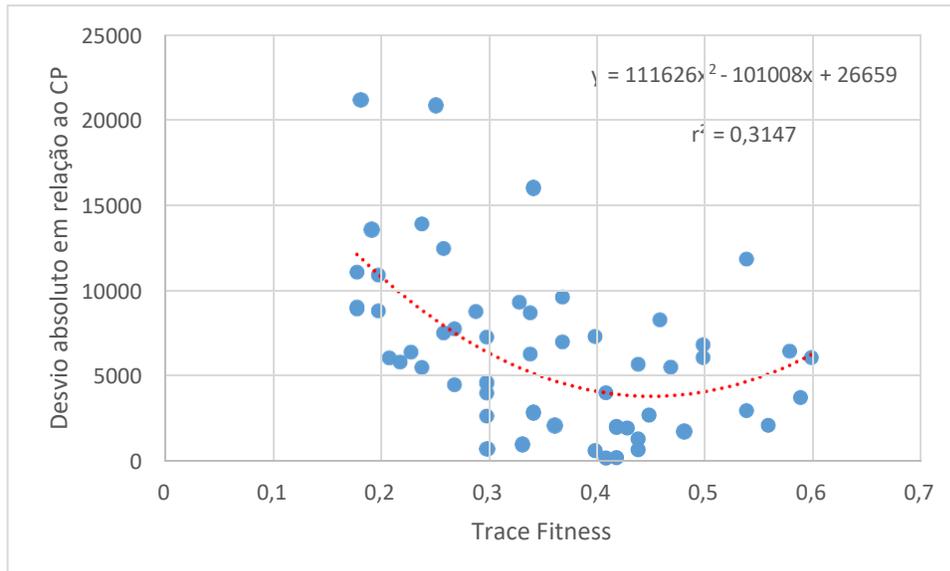


Ilustração 4.5 Diagrama de dispersão entre o trace fitness e o desvio absoluto em relação ao CP

Através da observação dos diagramas é observa-se que, em ambos os casos, foram utilizadas regressões polinomiais de segundo grau para tentar explicar as variáveis custo total e desvio absoluto em relação ao CP com recurso à variável *trace fitness*. Analisa-se o coeficiente de determinação (r^2) e verifica-se que ambas as regressões têm baixa qualidade (r^2 a rondar o valor 0,3). Observa-se que a linha de tendência em ambos os casos tem na maior parte do intervalo um declive negativo o que indica que maiores níveis

de *trace fitness* estão normalmente associados a custos ou desvios mais reduzidos. No entanto é arriscado avaliar esta relação com muita certeza porque o valor do coeficiente de determinação mostra que a variável dependente não está a ser totalmente explicada pela regressão polinomial usando o *trace fitness* como variável independente. O baixo valor do coeficiente de determinação pode estar relacionado com um dos seguintes fatores que podem ter influenciado a precisão da análise:

(1) grau de agregação - analisando comparativamente os dois mapas de processos nos anexos 2 e 3 observamos que existe um elevado nível de agregação. O primeiro é basicamente o CP original fornecido pelo Hospital da Luz enquanto que o segundo foi o mapa utilizado na análise. Pode ter acontecido que um grau de agregação excessivo tenha prejudicado a análise;

(2) A qualidade dos dados também pode ter estado na génese de alguns problemas. Este argumento vem do facto de, por exemplo, a atividade “consulta”, presente na generalidade dos pacientes não ter um custo padrão e aparecer muitas vezes com valor 0€, o que não parece muito coerente com a tipologia da instituição (privada). Esta atividade conta com o intervalo de valor [0; 684,22]€, onde 92 dos 521 eventos (consulta) têm valor 0€ (quase 20% do total das atividades consulta) e em que a média é de 43,47€;

(3) O último facto que pode ter influenciado negativamente os resultados foi a afetação do custo dos registos retirados que não constituíam atividades. É explicado, no ponto 3.2, que durante o processamento dos dados foi necessário retirar registos que não constituíam atividades, mas que tinham custos associados (materiais, fármacos, etc.). Estes custos foram afetados a atividades no registo de cada doente, tendo-se usado critérios de proximidade temporal entre os registos. Assim, tentou-se afetar sempre custos referentes a registos retirados a atividades que tivessem a mesma especialidade. No entanto, nem sempre se verificou todos ou algum dos critérios, este facto pode ter causado algum desvio pode ter tido algum impacto nos resultados finais.

CAPÍTULO 5: DISCUSSÃO

Neste capítulo exploram-se os resultados obtidos, as suas implicações e a relação existente entre os mesmos. Verifica-se que existem atividades que, em termos de custo, são mais preponderantes e que são generalizadamente aplicadas, como cirurgia e internamento. Outras têm elevados custos associados, como a quimioterapia e a radioterapia, mas, contra o que seria expectável, são aplicadas unicamente a um reduzido número de casos, dentro da amostra utilizada. As restantes são atividades com custos menos elevados e cujo impacto tem sobretudo a ver com o número de vezes que são realizadas em cada doente, como, por exemplo, os métodos complementares de diagnóstico e tratamento (MCDT). Espera-se, de acordo com o que foi apresentado no enquadramento teórico, que seja possível refletir a eficiência de custo através dos níveis de conformidade para os processos estudados. Desse modo, foi estudada e verificada uma relação entre o custo total de cada doente e o número de atividades a que o mesmo foi sujeito. Para além disso, verificou-se também uma relação entre o desvio absoluto do custo total do doente em relação ao custo standard do CP e o número de atividades a que o utente foi sujeito. No primeiro caso, o coeficiente de correlação é igual a 0,6910, significando que existe uma correlação moderada no mesmo sentido entre o custo total e o número de atividades realizado. No segundo, o valor do mesmo coeficiente é de 0,5360 indicando também uma evolução moderada conjunta das duas variáveis (apresentados graficamente em anexo 1). Quanto às relações que são o ponto essencial deste estudo, ou seja, as relações entre os custos (totais e desvios) e o nível de conformidade conclui-se, tendo em primeiro lugar em atenção as representações gráficas, que existe uma relação fraca (r^2 baixo) quer entre o custo total e o *trace fitness* (nível de conformidade), quer entre o valor absoluto do desvio e o *trace fitness*. Nestes casos verificam-se relações inversas das variáveis, ou seja, verifica-se que existe uma tendência decrescente na regressão polinomial, que nos dizem que maior nível de conformidade está habitualmente associado a menores custos e, portanto, a menores desvios. A apoiar este facto surgem os coeficientes de correlação no caso dos custos totais em relação ao *trace fitness* com valor de -0,5785 e para a relação do valor absoluto dos desvios e o *trace fitness* de -0,5609. Estes coeficientes de correlação apoiam a hipótese da influência da conformidade na eficiência de custos, desenvolvida no enquadramento teórico, como foi já referida previamente neste ponto. No entanto, o coeficiente de determinação baixo indica que será preciso uma melhor qualidade de dados ou uma análise multivariada (com atenção aos *outcomes* dos doentes, ou de uma outra métrica de avaliação da eficácia dos

tratamentos) para ter de facto uma explicação totalmente fiável da eficiência dos custos de tratamento para o CCR no Hospital da Luz de Lisboa.

CAPÍTULO 6: CONCLUSÕES

Neste capítulo realiza-se a ligação entre os resultados e as perguntas de investigação para tentar inferir algumas conclusões. Em resposta à primeira pergunta de investigação (Qual o nível de conformidade entre o processo real e o processo standardizado?) observa-se, nos resultados obtidos através do ProM Tools que o nível médio de conformidade é de 0,3548, o que se pode entender como que existe correlação, entre o processo real e o processo standardizado, no entanto esta será de facto reduzida. Existem várias explicações possíveis para este nível de conformidade não ser mais elevado. Em primeiro lugar, e como já foi referido anteriormente, no sector da saúde, em particular quando se tratam de patologias mais agressivas e cujo tratamento envolve mais etapas, cada paciente vai ter a sua própria reação aos tratamentos e nem todos os tratamentos que são tidos como eficazes o serão para todos os pacientes. Adicionalmente, neste estudo é focado exclusivamente o tratamento do CCR, contudo outro dos motivos que influenciam a conformidade são a existência de co-morbilidade, por outras palavras, de outras complicações que o utente possua ou desenvolva durante o tratamento desta patologia e que enfraquecem o seu estado de saúde. Alternativamente, existe também a possibilidade de a responsabilidade pela falta de forte conformidade com o processo standardizado seja dos responsáveis clínicos. Cabana et al. (1999) descrevem motivos que levam os médicos a não seguirem os CP. Foram identificados sete tipos de barreiras à aplicação de processos que afetam o conhecimento, as atitudes e o comportamento dos profissionais de saúde. As sete barreiras identificadas foram a falta de conhecimento e a falta de familiaridade (impactantes no conhecimento), a falta de consenso, falta de autoeficácia, a falta de expectativas em relação aos resultados e a inércia de procedimentos anteriores (impactantes nas atitudes) e barreiras externas (impactantes no comportamento). Algumas destas barreiras são bastante autoexplicativas, no entanto outras podem parecer estranhas ao leitor e são descritas da seguinte forma. A falta de consenso prende-se com a não concordância dos clínicos em relação às diretrizes ou a algum tipo de comportamento padrão; a falta de autoeficácia prende-se com a convicção na capacidade de seguir um determinado procedimento; a inércia de procedimentos anteriores quer dizer que os clínicos estão

automatizados para um determinado comportamento e não têm a motivação necessária para a mudança; e as barreiras externas, segundo Cabana et al. (1999) podem ter três tipologias distintas: relacionadas com as diretrizes em si, com o doente ou com o ambiente. Relacionadas com as indicações as barreiras costumam ser a inconveniência ou dificuldade de aplicação; relacionadas com os doentes estão ligadas à dificuldade de coincidir as preferências do utente com o CP; e a relação com o ambiente são descritas como estando fora do controlo do clínico e associadas às infraestruturas ou à aquisição de equipamento específico.

Passando agora para a segunda pergunta de investigação (Qual o impacto financeiro da variação entre processos reais e estandardizados?) os resultados apontam para um real impacto financeiro da variação entre os processos reais e o processo estandardizado. Isto pode ser visto através da relação anteriormente descrita entre as variáveis *trace fitness* e custo total ou alternativamente com o valor absoluto dos desvios em relação ao CP. O coeficiente de correlação negativo para as duas relações indica que um maior desvio do CP implica um menor custo. Selecionando os quatro casos limite, ou seja, os quatro casos com menor *trace fitness* e os quatro casos com maior valor desse campo e fazendo a média do valor dos desvios obtemos que no caso de baixa conformidade o valor médio dos desvios é de 12.511,08€, enquanto, para o caso oposto o valor médio dos desvios é de 4.593,87€. Comparando estes valores a diferença é de 7.917,20€ o que parece indicar um elevado nível de impacto financeiro associado à variação em relação ao CP. De forma a obter um outro modo de comparação fazendo a média dos desvios para os casos com *trace fitness* mais próximo ao valor global obtém-se a média do valor absoluto dos desvios em relação ao CP de 8.384,43€ validando o que os outros resultados indicavam. Desta forma e, tendo também presente o que foi dito no ponto anterior, é possível dizer, com relativa segurança, que existe de facto um impacto financeiro relativo à conformidade com o processo standard. Todos os valores usados neste ponto são obtidos através da tabela em anexo 1.

No respeitante a trabalhos futuros relevantes, identificam-se necessidades a responder e possíveis análises a efetuar. Este relatório realizou-se em torno da base de dados do Hospital da Luz de Lisboa em relação a utentes diagnosticados com CCR nos anos de 2017 e 2018. O tamanho da amostra inicial é reduzido (116) e com o processamento necessário ainda se torna mais pequeno (56). Uma das recomendações para a realização de trabalhos futuros prende-se exatamente com esta limitação. Num mundo cada vez mais informatizado, a integração da informação e o crescente poder de armazenamento

de dados permitem que os registos sejam cada vez mais detalhados e, dessa forma, uma possível amostra usada num futuro estudo seja mais completa. Quando se recorre à mineração de processos, a descoberta de padrões é mais fiável com uma base de dados mais extensa. Para além disso, seria interessante tirar partido de uma equipa multidisciplinar para conseguir fazer um paralelismo com o tema tratado neste relatório e, por exemplo, acrescentar os resultados de cada paciente a cada procedimento para haver uma crescente aproximação ao conceito de *value based healthcare* que, de certa forma, motivou o hospital a criar esta oportunidade de estágio curricular. Para além disso, seria também interessante utilizar outras vertentes de *process mining* para analisar os *event logs* produzidos pela instituição de saúde. Para isso, mais uma vez, seria benéfica a integração dos registos de todo o grupo Luz Saúde para uma determinada patologia para poder, além de fazer a análise geral, fazer também a análise em termos comparativos entre as diferentes instituições do grupo ou até por regiões. Por último, seria extremamente benéfico para futuros trabalhos nesta área que o registo do hospital funcionasse através de uma outra ferramenta, por exemplo SAP, que permite não só um registo mais eficiente e completo como exportação direta do mesmo em formato de *event log*.

Bibliografia

1. Adriansyah, A.; van Dongen, B. F.; van der Aalst, W.M.P. (2011): Conformance Checking Using Cost-Based Fitness Analysis.; proceedings. 2011 15th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2011). Helsinki, Finland, 8/29/2011 - 9/2/2011. Annual IEEE Computer Conference; IEEE International Enterprise Distributed Object Computing Conference; EDOC; pág. 55–64.
2. Barros, S. (2018) ; Conformance Checking for Care Pathway Compliance Assessment (Master Thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa)
3. Cabana, M. D.; Rand, C. S.; Powe, N. R.; Wu, A. W.; Wilson, M. H.; Abboud, P. A.; Rubin, H. R. (1999): Why don't physicians follow clinical practice guidelines? A framework for improvement. Em: *JAMA* 282 (15), pág. 1458–1465. DOI: 10.1001/jama.282.15.1458.
4. Canensi, L; Leonardi, G; Montani, S; Terenzi, P (2018): A Context-Aware Miner for Medical Processes. Journal of e-Learning and Knowledge Society, Vol 14 No 1 (2018): Journal of e-Learning and Knowledge Society - Focus on: New trends, challenges and perspectives on healthcare cognitive computing: from information extraction to healthcare analytics. DOI: 10.20368/1971-8829/1453.
5. Clerc, L.; Jooste, V.; Lejeune, C.; Schmitt, B.; Arveux, P.; Quantin, C. et al. (2008): Cost of care of colorectal cancers according to health care patterns and stage at diagnosis in France. Em: *The European journal of health economics : HEPAC : health economics in prevention and care* 9 (4), pág. 361–367. DOI: 10.1007/s10198-007-0083-0.
6. Fei, H.; Meskens, N. (2010): Discovering patient care models from event logs. 8th International Conference of Modeling and Simulation - MOSIM'10 - May 10-12, 2010 - Hammamet - Tunisia.
7. Hoverman, J.; Cartwright, T.; Patt, D.; Espirito, J.; Clayton, Matthew P.; Garey, Jody S. et al. (2011): Pathways, outcomes, and costs in colon cancer: retrospective evaluations in 2 distinct databases. Em: *The American journal of managed care* 17 Suppl 5 Developing, SP45-52.
8. Huang, Z.; Bao, Y.; Dong, W.; Lu, X.; Duan, H. (2014): Online treatment compliance checking for clinical pathways. Em: *Journal of medical systems* 38 (10), pág. 123. DOI: 10.1007/s10916-014-0123-0.
9. Mans, R.; Schonenberg, H.; Leonardi, G.; Panzarasa, S.; Cavallini, A.; Quaglini, S.; van der Aalst, W. (2008): Process mining techniques: an application to stroke care. Em: *Studies in health technology and informatics* 136, pág. 573–578.
10. Mar, J.; Errasti, J.; Soto-Gordoa, M.; Mar-Barrutia, G.; Martinez-Llorente, J.; Domínguez, S. et al. (2017): Valoración del coste económico del cáncer colorrectal según estadio tumoral. Em: *Cirugía española* 95 (2), pág. 89–96. DOI: 10.1016/j.ciresp.2017.01.001.
11. Milchak, J. L.; Carter, B. L.; Ardery, G.; Black, H. R.; Bakris, G. L.; Jones, D. W.; Kreiter, C. D. (2006): Development of explicit criteria to measure adherence to hypertension guidelines. Em: *Journal of human hypertension* 20 (6), pág. 426–433. DOI: 10.1038/sj.jhh.1002005.
12. Müller, R.; Rogge-Solti, A. (2011): BPMN for Healthcare Processes.
13. Rojo, M.; Rolón, E.; Calahorra, L.; García, F.; Sánchez, R.; Ruiz, F. et al. (2008): Implementation of the Business Process Modelling Notation (BPMN) in the modelling of anatomic pathology processes. Em: *Diagnostic pathology* 3 Suppl 1, S22. DOI: 10.1186/1746-1596-3-S1-S22.
14. Rovani, M.; Maggi, F.; Leoni, M.; van der Aalst, W. M.P. (2015): Declarative process mining in healthcare. Em: *Expert Systems with Applications* 42 (23), pág. 9236–9251. DOI: 10.1016/j.eswa.2015.07.040.
15. Van de Steeg (2015): Process Mining. Mining for costs and (near) incidents. Master thesis, Eindhoven University of Technology, The Netherlands

- 16 Van der Aalst, W.; Adriansyah, A.; Medeiros, A.; Arcieri, F.; Baier, T.; Blickle, T. et al. (op. 2012): Process Mining Manifesto. Em: Florian Daniel, Kamel Barkaoui e Schahram Dustdar (Ed.): Business process management workshop (BPD 2011). Revised selected papers, vol. 99. Heidelberg: Springer (Lecture notes in business information processing (Internet), vol. 99), pág. 169–194.
- 17 Van der Aalst, W.; van Dongen, B.; Herbst, J.; Maruster, L.; Schimm, G.; Weijters, A. (2003): Workflow mining: A survey of issues and approaches. Em: *Data & Knowledge Engineering* 47 (2), pág. 237–267. DOI: 10.1016/S0169-023X(03)00066-1.
- 18 Weidlich, M.; Polyvyanyy, A.; Desai, N.; Mendling, J.; Weske, M. (2011): Process compliance analysis based on behavioural profiles. Em: *Information Systems* 36 (7), pág. 1009–1025. DOI: 10.1016/j.is.2011.04.002.
- 19 Weske, M. (2012): Business process management. Concepts, languages, architectures. 2. ed. Berlin: Springer (Computer science).
- 20 White, S. (2004): Introduction to BPMN. IBM Corporation.



ANEXO 1 – TABELA DE DADOS E GRÁFICOS DE DISPERSÃO

Tabela 1- Dados de pacientes e atividades

Case	Actos de Enfermag	Cirurgia	Consulta	Imunohemoterapia	internamento	MCDTs	MCDTs Especialidades	MCDTs Especialidades - Gastroenterologia	MCDTs Imagiologi	Quimioterapia adjuvante	Quimioterapia neo-adjuvante	Radioterapia	Total	Nº atividade	Trace Fitness	Desvio
110312279	0	15231,85	775,25	349,31	8764,65	1520,97	302,45	423,45	912,14	570,35	503,32	0	29353,74	89	0,18	21060,58
106999311	0	6175,8	436,55	0	8560,86	1475,48	35,45	0	629,63	0	0	0	(17313,77)	86	0,18	9020,61
104349557	48,3	6777,78	1578,92	38	5527,2	898,04	1177,6	243,8	810,75	0	0	0	0 17100,39	85	0,2	8807,23
104581794	0	7391,2	707,25	138,62	5741,24	1293,75	565,8	1060,66	2291,02	0	0	0	(19189,54)	85	0,2	10896,38
110730917	0	4979,11	243,7	1366,66	9207,76	594,57	93,87	161,49	555,29	0	0	0	0 17202,45	83	0,18	8909,29
100244530	0	9093,29	148,61	381,22	8060,37	276,33	1194,54	0	192,63	0	0	0	(19346,95)	74	0,18	11053,83
110328310	29,9	14754,37	106,15	0	2956,12	356,46	93,87	0	701,54	0	0	0	2753,1 21751,51	69	0,19	13458,35
101881916	0	6538,99	723,35	0	1132,12	1714,1	781,48	0	3797,32	0	0	0	(14687,36)	65	0,23	6394,2
100522718	15	7620,39	836,6	0	2789,51	887,11	800,95	241,5	1161,68	0	0	0	0 14352,74	62	0,21	6059,58
106331299	0	0	970,11	601,74	3684,28	993,91	120,32	1173,52	388,29	0	5851,6	0	(13783,77)	60	0,24	5490,61
106847031	56,35	7506	1625,24	0	1419,34	511,89	46	0	701,5	2238,59	0	0	0 14104,91	59	0,22	5811,75
101425163	0	12997,5	655,52	1090,15	3415,13	660	736	244,03	943,01	0	0	0	(20741,34)	55	0,26	12448,18
100717764	0	5610,04	948,18	0	1941,05	606,72	93,87	479,35	798,63	460	0	0	0 10937,84	53	0,3	2644,68
110653262	0	11275,31	414	0	8770,89	679,79	459,08	0	569,2	0	0	0	(22168,27)	53	0,24	13875,11
102158185	0	11719,48	268,63	138,62	2229,8	363,97	147,14	150,31	1028,31	0	0	0	0 16046,26	52	0,27	7753,1
112012671	0	17561,06	150,09	0	9679,44	1094,24	175,08	0	393,88	0	0	0	(29053,75)	51	0,25	20760,63
110701860	0	7910,24	474,4	0	904,06	409,78	298,43	330,63	230	4997,35	0	0	0 15554,89	47	0,3	7261,73
104781009	167,64	9146,35	410,55	0	1949,06	582,07	115	0	399,63	0	0	0	(12770,3	46	0,27	4477,14
108611825	0	10318,88	445,05	331,49	3073,08	338,1	766,14	626,43	1074,7	0	0	0	0 16973,87	43	0,34	8680,71
107530520	0	9991,33	540,55	0	5012,65	829,58	115	0	575	0	0	0	(17064,11)	41	0,25	8770,95
100169716	0	8134,2	442,75	138,62	1766,09	673,08	417,45	517,79	833,75	0	0	0	0 12923,73	40	0,3	4630,57
108125882	197,8	19612,28	169,84	0	2892,43	94,99	0	404,6	812,62	0	0	0	(24184,56)	40	0,34	15891,4
102656029	0	0	209,88	0	0	1045,19	374,76	345	1766,43	0	0	0	0 3741,26	36	0,3	4551,9
102953228	33,8	0	621,99	38	2816,52	195,27	276	80,5	230	0	0	0	(4292,06	36	0,3	4001,08
109314000	12,1	0	204,7	331,49	1914,09	516,29	350,75	857,13	357,31	2931,81	0	0	0 7475,67	36	0,33	817,49
110632168	31,78	9816,72	106,15	0	692,93	337,42	153,53	404,6	628,89	3102,04	0	0	(15274,06)	36	0,37	6980,9
107158914	36,25	3687,74	139,47	0	4436,62	208,71	34,07	241,5	205,38	0	0	0	0 8989,74	34	0,3	696,58
111184070	0	0	1894,81	349,31	1398,03	657,89	230	386,4	694,31	0	0	0	(5610,75	34	0,34	2682,41
107332450	92	2968,57	190,9	331,49	1332,78	223,06	506	1493,72	575	0	0	0	0 7713,52	32	0,4	579,64
109482182	9,67	6883,85	106,15	0	1469,2	211,89	399,47	163,3	986,38	0	0	0	(10229,91)	32	0,36	1936,75
105322300	89,7	0	569,86	0	0	137,34	241,5	273,7	697,85	0	0	0	0 2009,95	31	0,34	6283,21
109105184	0	15224,24	130,53	0	0	1039,83	345	138	1025,61	0	0	0	(17903,21)	31	0,37	9610,05
100483569	0	7541,45	148,61	0	765,18	275,15	732,71	404,6	332,68	0	0	0	0 10200,38	30	0,43	1907,22
110382366	89,7	8129,21	294,4	0	1485,37	242,22	115	0	362,25	0	0	0	6873,72 17591,87	29	0,33	9298,71
102765693	0	14270,74	571,37	138,62	816,46	387,22	34,07	150,31	203,43	0	0	0	0 16572,22	27	0,46	8279,06
102739471	0	5933,17	258,15	0	1745,06	86,53	57,74	0	375,63	0	0	0	(8456,26	26	0,42	163,12
100358411	0	6476,23	199,16	0	880,87	162,52	401,53	665,5	787,76	0	0	0	0 9573,57	24	0,44	1280,41
110338120	0	0	285	0	935,66	424,92	287,5	0	853,59	0	0	0	(2786,67	22	0,47	5506,49
111710103	0	7143,47	296,25	0	970,95	93,38	516,35	94,35	345	2574,66	0	0	0 12034,41	22	0,59	3741,25
109326474	6,35	7090,59	91,27	601,74	1428,13	342,24	34,07	0	282,63	0	0	0	(9877,02	21	0,48	1583,86
107290898	0	7777,1	657,04	0	1386,86	573,79	230	0	633,94	0	0	0	0 11258,73	18	0,54	2965,57
110359143	0	12765,3	42,46	6,06	6378,64	637,45	0	0	294,13	0	0	0	(20124,04)	18	0,54	11830,88
105631752	0	9149,24	690,3	201,91	1562,64	414,65	287,5	0	0	0	0	0	0 12306,24	17	0,41	4013,08
110452534	28,63	0	183,08	0	596,42	206,38	589,59	1005,12	0	0	0	0	(2609,22	17	0,44	5683,94
110483510	0	6753,29	640,87	0	1056,38	111,51	316,25	1535,74	0	0	0	0	0 10414,04	17	0,56	2120,88
102311277	22,95	7661,29	184	0	1024,34	7,82	355,82	696	279,65	0	0	0	(10231,87)	16	0,42	1938,71
102543763	0	0	63,69	0	854,34	230,69	93,87	322,61	282,63	0	0	0	0 1847,83	16	0,58	9445,33
110295005	0	6376,42	120,75	0	765,96	215,01	46	0	126,5	0	0	0	(7650,64	15	0,44	642,52
110730194	0	0	155,09	0	1073,82	589,87	0	0	393,88	0	0	0	0 2212,66	14	0,5	6080,5
111725470	0	4855,21	98,28	0	1095,81	38,65	34,07	0	172,5	0	0	0	(6294,52	14	0,42	1998,64
107184958	0	0	0	0	0	82,88	108,82	0	595,59	0	0	0	0 787,29	13	0,26	7505,87
104817763	0	0	0	0	0	94,3	229,54	345	317,03	0	0	0	(985,87	10	0,4	7307,29
106589580	0	0	80,5	0	0	303,83	1147,24	373,75	319,82	0	0	0	0 2225,14	10	0,6	6068,02
109496744	0	0	120,75	0	632,61	25,08	230	0	460	0	0	0	(1468,44	10	0,5	6824,72
110694848	0	4396,51	120,75	0	776,31	0	315,68	0	15,15	0	0	0	0 5624,4	10	0,45	2668,76
103335293	0	7619,71	100,63	0	717,58	0	0	0	0	0	0	0	(8437,92	7	0,41	144,76
Total	967,92	372865,5	22648,13	6573,05	140486,69	26973,91	17609,95	16034,39	34401,47	16874,8	6354,92	9626,82	671417,6			
% total geral	0,144160664	55,534071	3,37318112	0,978980963	20,92389304	4,017457	2,622801564	2,388139839	5,123707297	2,513309341	0,946492983	1,433805238				
zeros	39	14	2	38	6	2	4	23	4	49	54	54				
Média	56,93647059	8877,75	419,409815	365,1694444	2809,7338	499,5169	338,6528846	485,8906061	661,5667308	2410,685714	3177,46	4813,41			0,35482143	

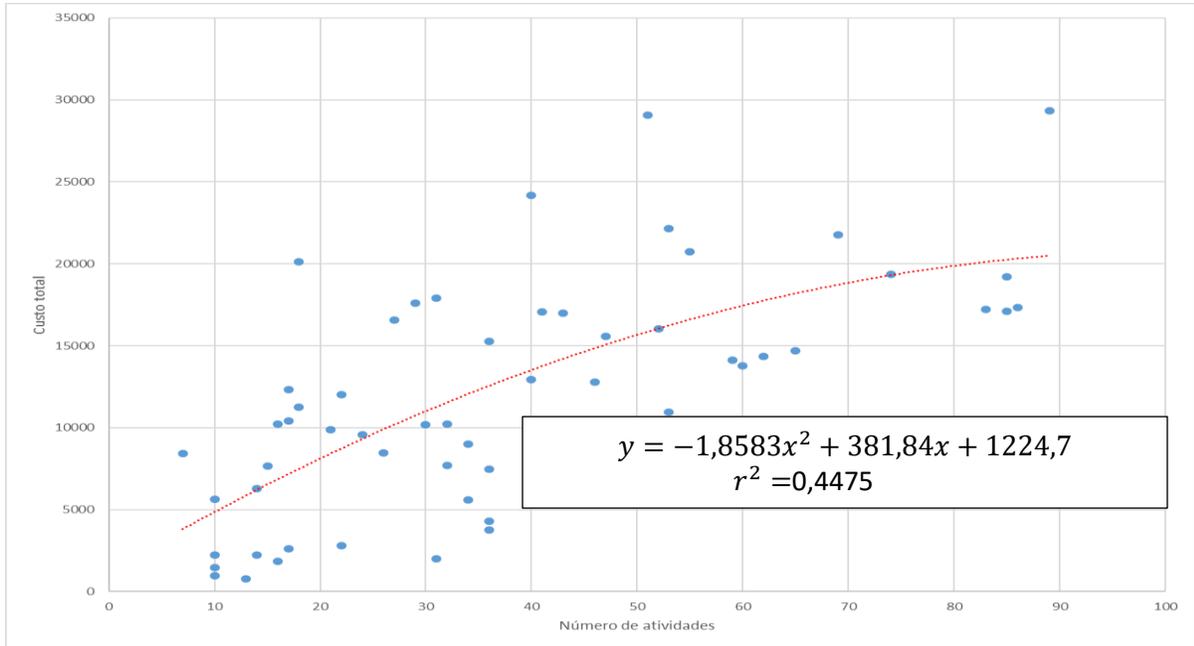


Figura 2 Gráfico de dispersão entre número de atividades e custo total

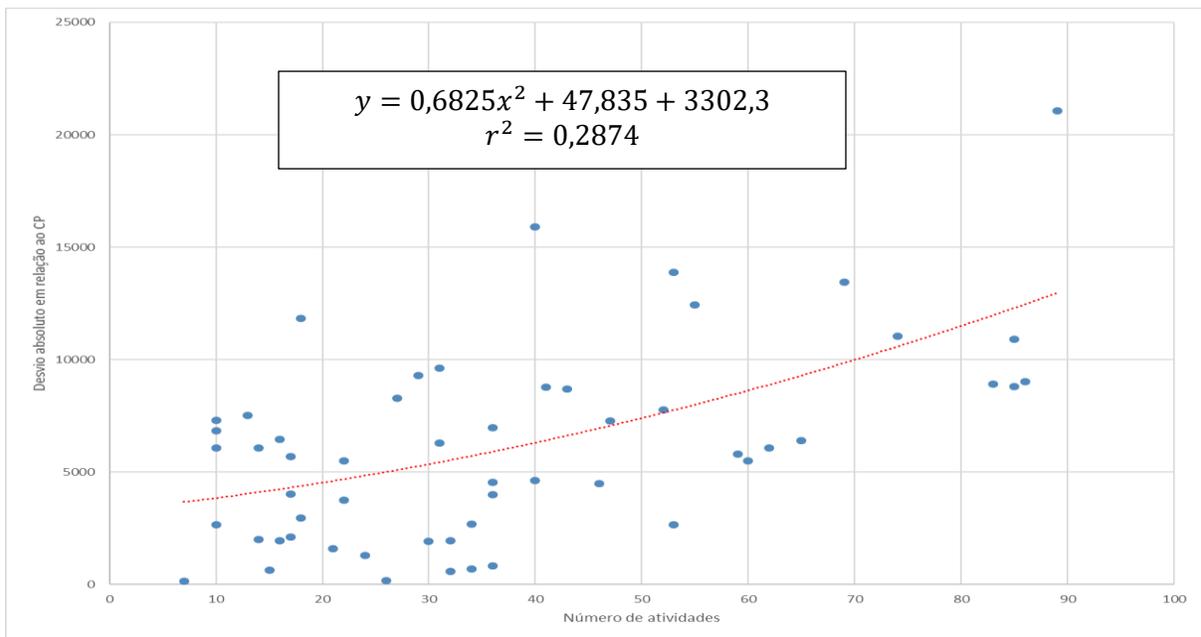


Figura 3 Gráfico de dispersão entre número de atividades e desvio absoluto em relação ao CP

ANEXO 2 – MODELO BPMN DO PROCESSO DE TRATAMENTO DO CCR SEGUNDO O HOSPITAL DA LUZ DE LISBOA

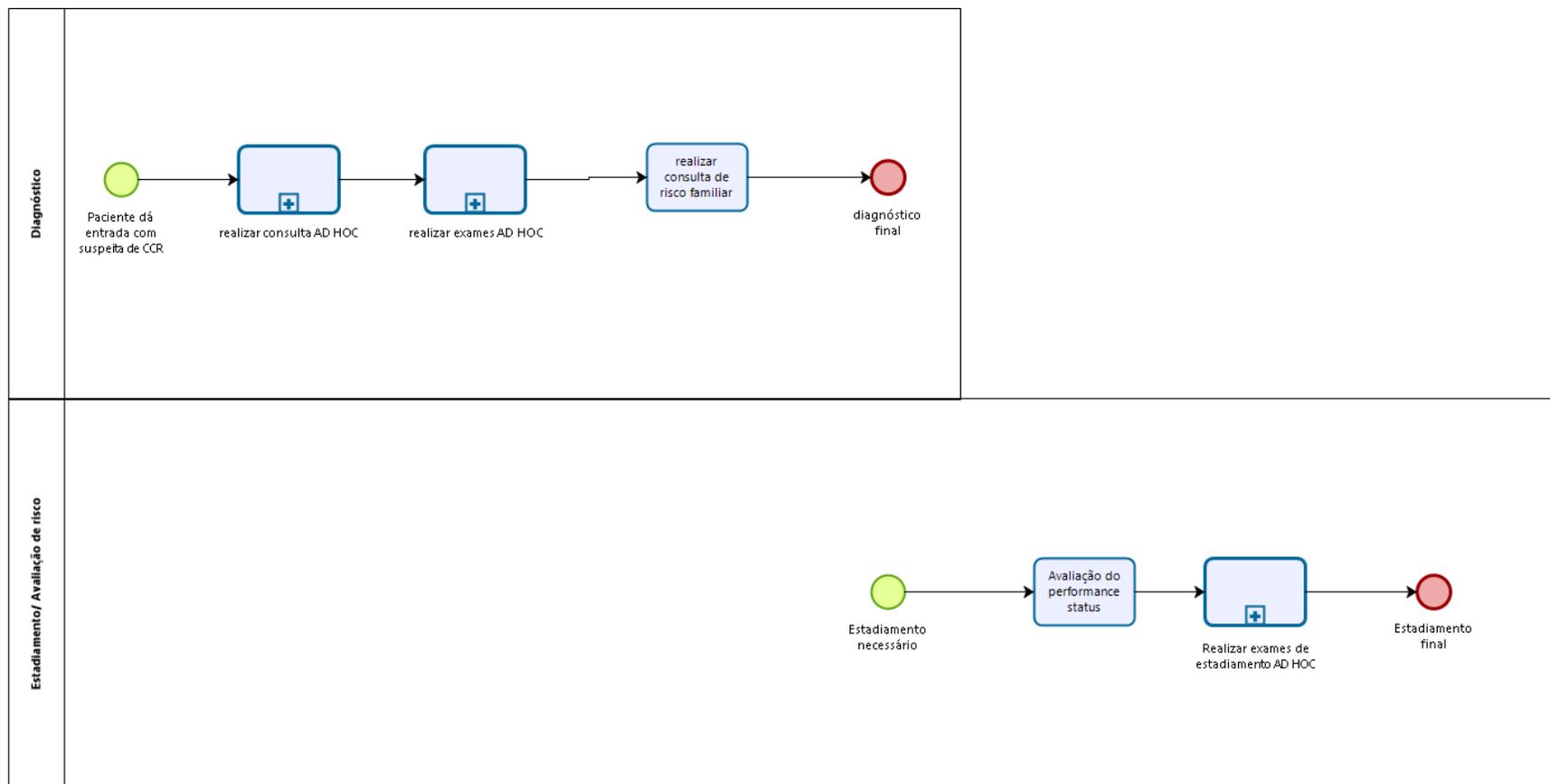


Figura 4 Parte 1 do modelo BPMN para o tratamento do CCR segundo o Hospital da Luz

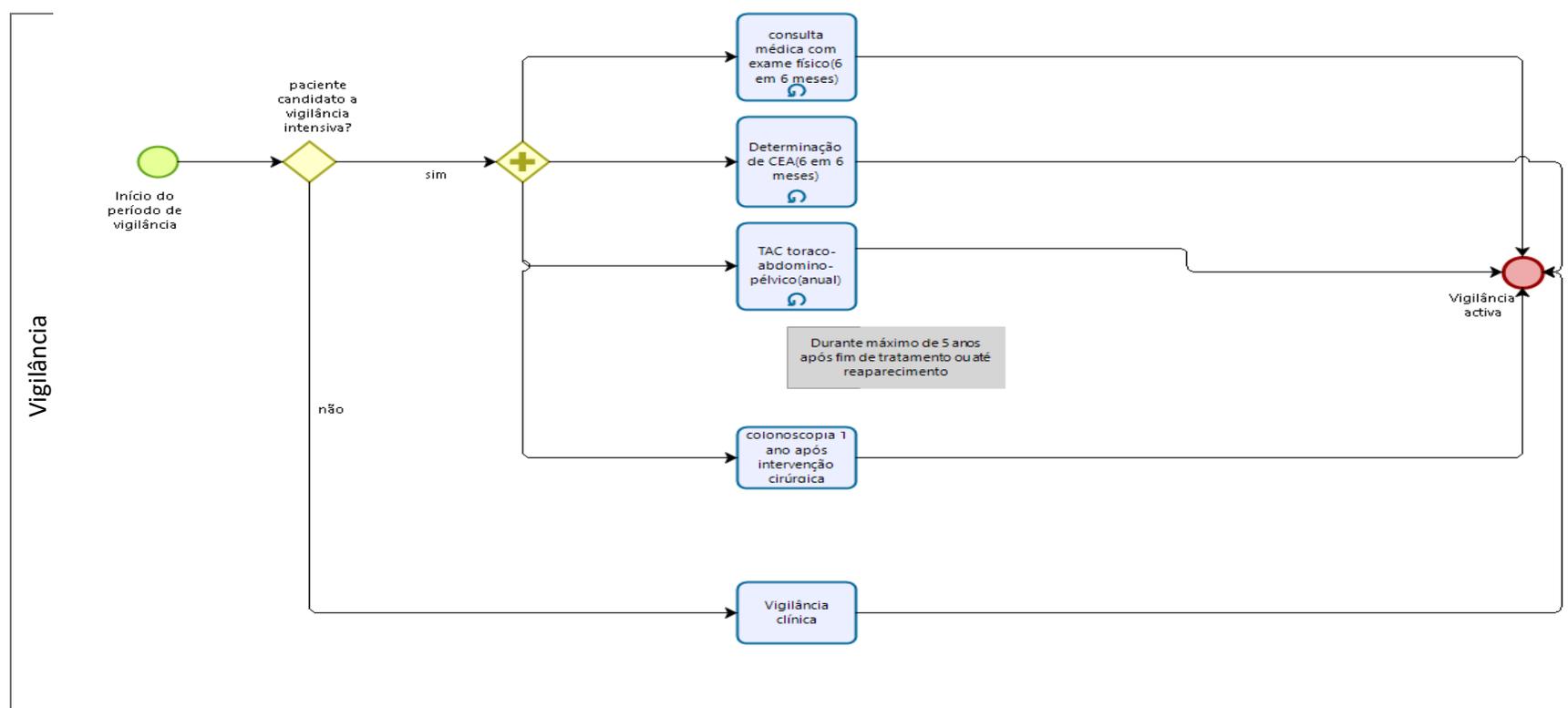


Figura 6 Parte 3 do modelo BPMN para o tratamento do CCR segundo o Hospital da Luz

ANEXO 3 – MODELO BPMN PARA O TRATAMENTO DO CCR ADAPTADO

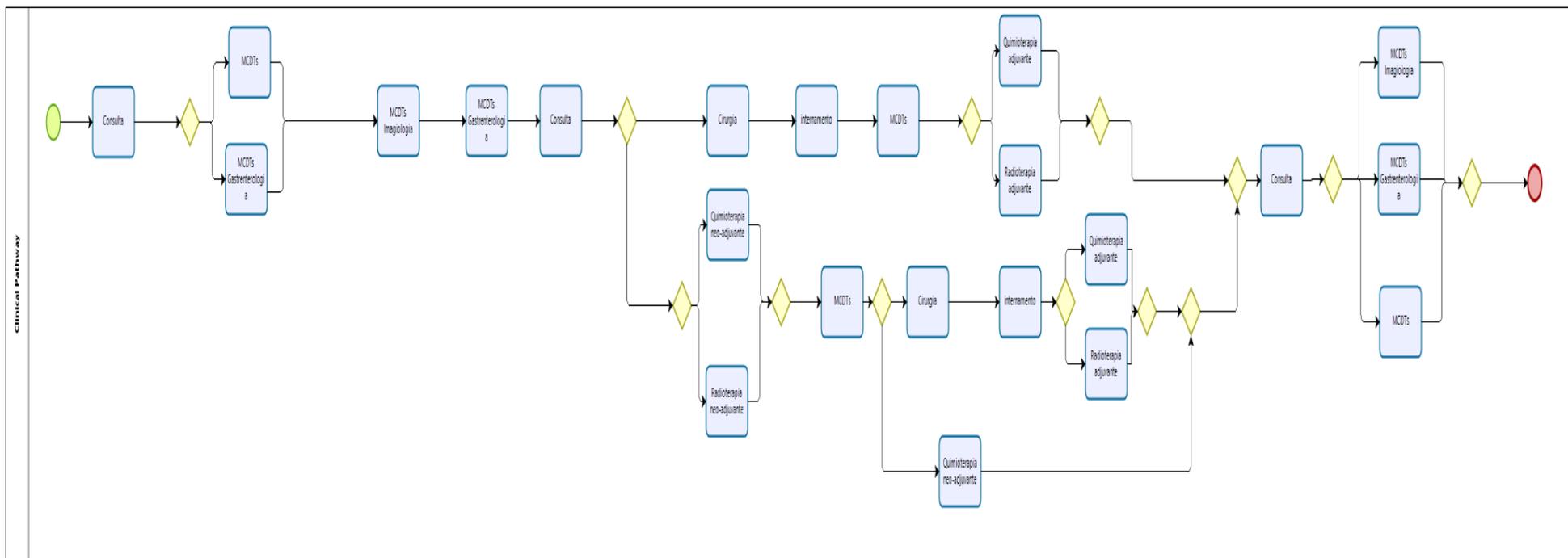


Figura 7 Modelo BPMN para o tratamento do CCR adaptado