

APLICAÇÃO DE UM ESTUDO DE SIMULAÇÃO EM UM CONTACT CENTER HOSPITALAR

Felipe Martins Pedrosa

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
fmpedrosa@hotmail.com

Wilkson Ricardo Silva Castro

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
wilkson.rsc@gmail.com

Arthur Harry Frederico Ribeiro Kramer

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
arthur.kramer@ct.ufrn.br

Ricardo Pires de Souza

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
ricardopires@ct.ufrn.br

Claudia Aparecida Cavalheiro Francisco

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
claudia@ct.ufrn.br

Eric Lucas dos Santos Cabral

Universidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova, CEP 59078-970 | Natal/RN - Brasil
ericlucascabral94@gmail.com

RESUMO

Centrais de teleatendimento hospitalares tem se tornado um fator essencial da experiência do paciente nas instituições médicas, entretanto, as demandas concorrentes pela prestação do serviço virtual incorre na geração de filas e tempos de esperas elevados, que vem a prejudicar a satisfação dos pacientes. Diante disto, o estudo objetiva elaborar um modelo de simulação do setor de teleatendimento de um hospital oncológico de grande porte afim de estudar o dimensionamento da força de trabalho. As medidas propostas, com base no modelo de simulação, incorrem em melhorias significativas em relação ao cenário atual, reduzindo o tempo de espera nos chamados de baixa complexidade e a taxa de abandono.

PALAVRAS CHAVE: Call Center Hospitalar; Simulação; Teoria das Filas.

Tópicos: SIM – Simulação, SA – PO na Área da Saúde.

ABSTRACT

Hospital call centers have become an essential factor of the patient's experience in medical institutions, however, competing demands for the provision of virtual service incurs in the generation of queues and long waiting times, which comes to impair patient satisfaction. The study aims to develop a simulation model for the telemarketing sector of a large oncology hospital in order to study the dimensioning of the workforce. The proposed measures, based on the simulation model, make significant improvements in relation to the current scenario, reducing the waiting time for calls with low complexity and the abandonment rate.

KEYWORDS: Medical Call Center; Simulation; Queueing Theory.

Paper topics: SIM – Simulation, SA – OR in Healthcare.

1. Introdução

Centrais de teleatendimento hospitalares tem se tornado um fator essencial da experiência do paciente nas instituições médicas, pois a ele cabe a função de ser a “porta de entrada virtual do paciente”, além de ter a autonomia para fornecer suporte contínuo durante toda a jornada na instituição, desde o encaminhamento personalizado antecedente ao serviço clínico até o aconselhamento individual após alta. Esta estratégia reduz o risco de readmissões desnecessárias e solidifica a lealdade do paciente [Stier, 1999; Owens e Stier, 2016]. Entretanto, as demandas concorrentes pela prestação do serviço virtual incorre na geração de filas e tempos de esperas elevados, que vem a prejudicar a satisfação dos pacientes. Por outro lado, os custos de eliminar as filas completamente são inviáveis, este dilema conhecido é o tópico principal do campo de estudo da teoria das filas.

O estudo proponente de teoria das filas data da primeira década do século XX, realizado pelo Dinamarquês Agner Krarup Erlang que aplicou conceitos de teoria da probabilidade a fim de resolver o problema de dimensionamento de operadores de linhas telefônicas com o objetivo de suprir a demanda de ligações na cidade de Conpenhague [Breuer e Baum, 2005; Lamarca, 2008]. Desde então muitos estudos se sucederam, sendo aplicados, por exemplo, para a solução de problemas de reparo de maquinário, problemas de escoamento de fluxo de barcos em terminais portuários, escoamento de pacientes de emergência em um hospital, entre tantas outras aplicações [Fogliatti e Mattos, 2013].

A Teoria das Filas é o estudo da espera em todas as formas diversas. Basicamente quaisquer sistemas nos quais as chegadas de entidades geram demanda para um recurso de capacidade limitada, pode ser chamado de um sistema de filas. Em particular se os tempos de chegada desta demanda forem imprevisíveis de forma que o conflito pela utilização do recurso eventualmente crescerá, ocasionando a formação de filas [Kleinrock, 1975; Hillier e Lieberman, 2013]. O estudo do sistema, dos elementos que o compõe e de suas medidas de desempenho facilitam a tomada de decisão quanto a manutenção ou mudança da sua infraestrutura, do seu dimensionamento de recursos humanos, dos seus equipamentos e instalações, a fim de alcançar um melhor desempenho global da operação [Fogliatti e Mattos, 2007]

Neste contexto, a teoria das filas é comumente aplicada em células de *call centers*, por configurarem um sistema de filas clássico, tendo em vista que é constituído de uma série de recursos chave, como mão de obra treinada, computadores e equipamentos de telecomunicação, que capacitam a organização de suprir a demanda por serviços de atendimento telefônico. Sob o ponto de vista gerencial, existem uma série de decisões a serem tomadas pela administração de uma operação de teleatendimento, entre elas: quantos agentes devem ser treinados para uma certa atividade; como agendar as escalas, intervalos e pausas destes agentes; qual o volume de ligações que é esperado de cada tipo; qual o tempo de operação de um chamado; qual a ordem de prioridade de chamados deve ser utilizada. Por essa razão, estudo relacionados a *call centers* estão muito atrelados à gestão efetiva da força de trabalho (ou WFM do inglês *Workforce Management*), pois embora seja uma operação de alto grau de dependência tecnológica, 70% dos custos operacionais dos *call centers* estão ligados à recursos humanos. No Brasil, o setor emprega 1,5 milhões de trabalhadores formais, e em lugares como os Estados Unidos e o Reino Unido é estimado que 3% da força de trabalho nacional sejam ligadas à serviços de teleatendimentos. [Koole e Mandelbaum, 2002; Brown et al., 2005; Erdem e Alaçam, 2007; Frost e Sullivan, 2016]

Devido ao alto grau de incerteza ligado a estas decisões, uma grande variedade de modelos matemáticos e softwares de gestão surgiram para auxiliar os líderes de operações a tomarem deliberações conscientes quanto as suas operações [Koole e Mandelbaum, 2002]. Nesta mesma crescente, tornou-se cada vez mais proeminente a utilização de simulação para guiar estas decisões [Mehrotra e Fama, 2003].

Simulação consiste em técnicas que permitem transpor o comportamento de sistemas reais para uma realidade simulada com o auxílio computacional. Alguns sistemas são simples o bastante para serem solucionados via modelos matemáticos, entretanto, a maioria dos problemas

reais são complexos demais para permitirem uma solução analítica. Neste pretexto, a simulação é utilizada para avaliar o modelo numericamente através de uma interface virtual, de forma a estimar as verdadeiras características desejadas no modelo real. [Law e Kelton, 2014]

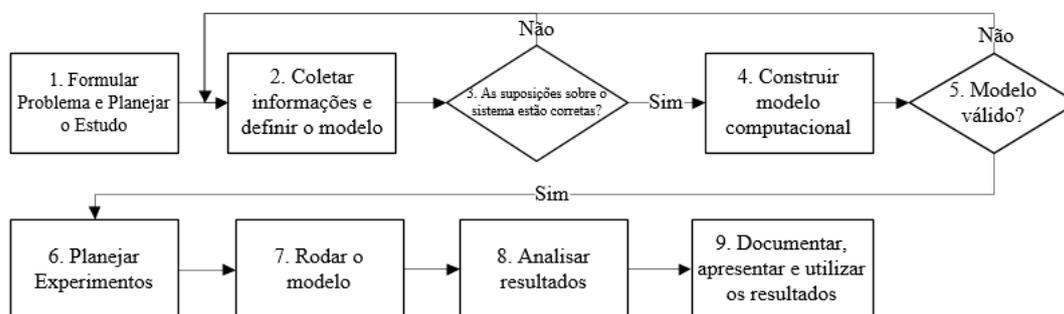
Bouzada (2009) destaca cinco fatores que contribuíram para a popularização da ferramenta de simulação para auxílio à gestão em *call centers*, entre eles: a crescente importância estratégica do call center para as organizações devido ao aumento de informações, dispositivos de comunicação e tecnologias, que tornou essencial a utilização de metodologias científicas em tomadas de decisões; a complexidade crescente no tráfego de chamadas juntamente com o roteamento de chamadas mais voltado para habilidades dos operadores; e os graus de incerteza cada vez maiores e mais predominantes acerca das decisões operacionais nos *call centers*.

Tendo por base estas considerações, o presente trabalho propôs a elaboração de um modelo de simulação computacional que busque simular o ambiente da central de teleatendimento de um centro oncológico, afim de dimensionar a força de trabalho de forma a aprimorar a performance do sistema, objetivando a diminuição dos tempos de espera para o atendimento e a taxa de abandono.

2. Metodologia e Estudo de Caso

A modelagem detalhada de uma simulação é parte de um esforço geral para entender e projetar um sistema complexo. Em razão disto, este estudo utilizou as etapas propostas por Law e Kelton [2014] para pautar seu percurso metodológico, conforme exibido no fluxograma da Figura 1. As Seções 2.2 e 2.3 a seguir detalham as atividades realizadas para a modelagem por meio de simulação do sistema investigado.

Figura 1. Passos para um estudo de simulação.



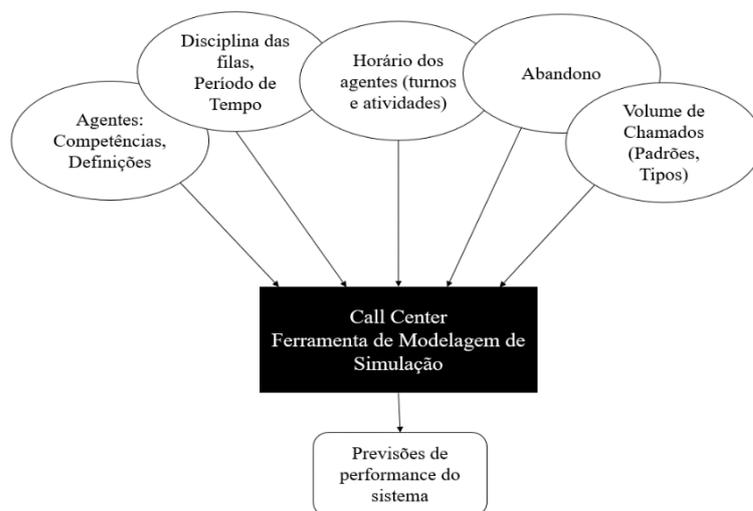
Fonte: Law [2014].

Na central de atendimento são realizados três processos: (i) Marcação de Consultas e Exames, (ii) Cancelamento ou Reagendamento, e (iii) Informação. Todos os processos supracitados podem ser operacionalizados através de dois meios: (i) pelo telefone, através do software de gerenciamento de chamadas, ou (ii) por mensagens digitais eletrônicas, que é gerenciado por outra plataforma de encaminhamento. Definiu-se como escopo deste estudo os atendimentos de origem por meio do telefone – englobando Marcação de Consultas e Exames e Informação - visto que representam a maior parte dos atendimentos. Portanto, desconsiderou-se os operadores que trabalham unicamente com o as telemensagens, assim como o operador responsável pelos reagendamentos.

2.2 Coletar Informações e definir o modelo

Para Mehtrotra e Fama (2003) o maior desafio na modelagem de simulação de um call center é a organização e separação dos inputs do modelo, visto que existem uma série de decisões a serem tomadas quanto ao nível de detalhe incluído, assim como a peculiaridade de que estes dados provém de múltiplas fontes diferentes, como pode ser observado na figura 2

Figura 2: Inputs necessários a uma simulação de *Call Center*.



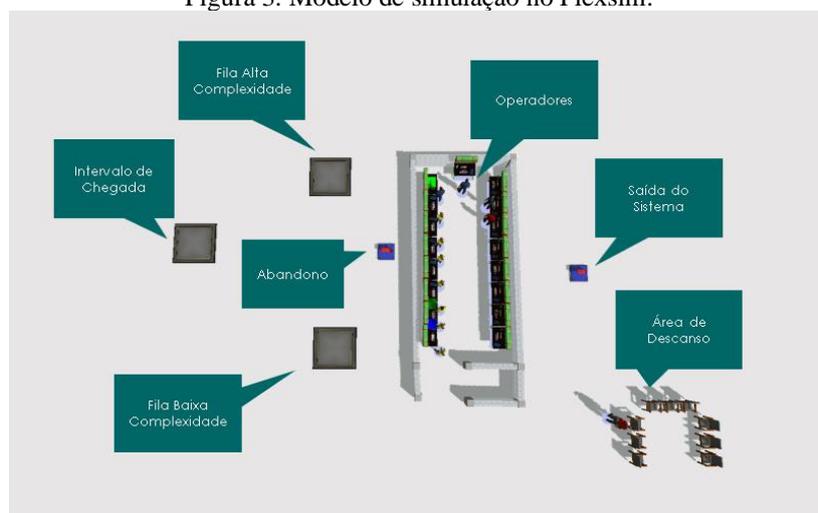
Fonte: Adaptado de Mehtrotra e Fama [2003]

Com relação aos dados disponíveis, o estudo teve acesso a extensa base de dados do software de gerenciamento de chamadas da unidade estudada, contando com 425.923 registros de chamados individuais gerados entre Janeiro de 2018 à Março de 2019. As informações quanto a escala e o treinamento dos agentes foram fornecidas pela gerência do setor de marcação através de planilhas eletrônicas. Tais dados foram analisados e tratados de modo a obter as informações relevantes para o processo de desenvolvimento do modelo de simulação.

2.3 Construção do modelo

O modelo de simulação proposto é composto, basicamente, pelos seguintes elementos: o intervalo de chegada, as filas de alta e baixa complexidade, o abandono, os operadores, a saída do sistema e a área de descanso, conforme esquematizado na Figura 3.

Figura 3. Modelo de simulação no Flexsim.



Fonte: Elaboração própria.

Os chamados chegam na caixa de intervalo de chegada dependentes da hora do dia do modelo, simulando a sazonalidade da vida real. Os tipos de chamados são definidos de acordo com as percentagens de frequência destes chamados observadas na base histórica de dados considerada.

Os chamados esperam em suas respectivas filas (alta complexidade ou baixa complexidade) e, conforme o seu tempo de espera aumenta, acresce sua probabilidade de

abandono. Desta forma, o chamado tem dois caminhos: (i) seguir para os operadores, ou (ii) abandonar a fila, indo neste caso para a caixa de “Abandono”.

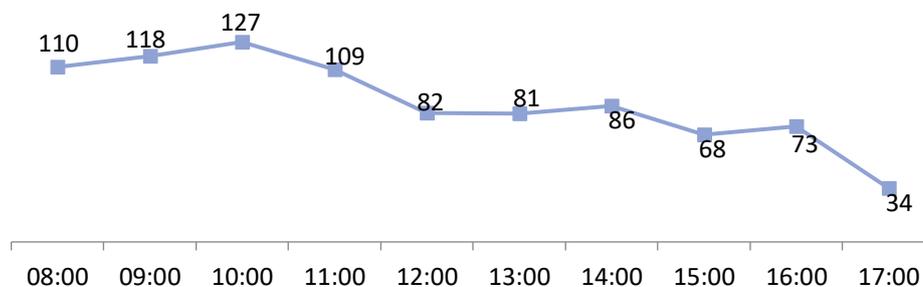
Caso os chamados sigam para os operadores, ele será escolhido por um operador conforme a ordem de prioridade adotada. Então, será operado com um tempo que variará dependendo de seu tipo (Consultas, por exemplo, demoram mais que Exames). Ao fim da operação, os chamados operados pelos atendentes seguem para a caixa de Saída do Sistema.

Durante o horário de trabalho os operadores tomam pausas, nestes períodos eles se dirigem para a “Área de descanso” e não operará nenhum chamado até que seu descanso acabe e ele retorne a sua estação de trabalho.

2.3.1 Geração de entidades

As chamadas apresentam um comportamento de sazonalidade diária, se concentrando pela manhã e tendo volume reduzido pela tarde, por esta razão optou por inserir distribuições de probabilidade conforme o horário do dia, conforme observado na Figura 4.

Figura 4. Média de chamados por hora do dia.



Fonte: Sistema de gerenciamento de chamadas.

As aderências as curvas de probabilidade foram realizadas através de extensão ExpertFit do *Flexsim* que utiliza o teste estatístico de Chi-Quadrado como padrão, e resultou nas distribuições apresentadas no Quadro 1.

Quadro 1. Distribuições de entrada.

Hora do Dia	Distribuição de probabilidade
8:00 - 9:00	Weibull[0.000000, 133.443328, 3.871688]
9:00 - 10:00	Weibull[0.000000, 162.835477, 4.554978]
10:00 - 11:00	Loglaplace[0.000000, 154.000000, 5.432672]
11:00 - 12:00	Weibull[0.000000, 142.611761, 4.452007]
12:00 - 13:00	Weibull[0.000000, 103.454506, 4.141867]
13:00 - 14:00	Loglogistic[0.000000, 89.771182, 6.325712]
14:00 - 15:00	Loglaplace[0.000000, 99.000000, 4.061985]
15:00 - 16:00	Weibull[0.000000, 88.663339, 3.409563]
16:00 - 17:00	Beta[0.145543, 237.923885, 8.112219, 13.956699]
17:00 - 18:00	Johnsonbounded[4.562764, 95.460152, 0.678884, 1.516908]

Fonte: Flexsim.

Em seguida o chamado gerado é classificado segundo a Unidade Remota de Atendimento (URA). Entre os tipos de chamados existem aqueles considerados de baixa e de alta complexidade. as marcações consideradas de baixa complexidade são aqueles referentes à: consulta, raios-X, ultrassonografia, mamografia, densitometria ou outros exames. Vide a simplicidade do procedimento, casos como estes requerem um menor grau de treinamento do operador. No que se refere às marcações de alta complexidade, nesta categoria estão inclusas: Ressonância Magnética, tomografia, PET-CT, medicina nuclear e cintilografia. Para a implementação deste comportamento

no modelo, utilizou-se a distribuição com base nos dados históricos de proporção entre chamados, indicado no Quadro 2.

Quadro 2. Classificação dos chamados e chance de ocorrência

Tipo de Item	Fila	Cor	Complexidade do Chamado	Chance de Ocorrência
1	Consulta	Vermelho	Baixa	42,3%
2	Exame	Verde	Baixa	26,5%
3	Oncologia Clínica	Azul	Baixa	8,2%
4	Radioterapia	Amarelo	Baixa	1,3%
5	Ambulatório Caicó	Laranja	Baixa	0,8%
6	Tomografia	Roxo	Alta	8,5%
7	Ressonância Magnética	Verde Limão	Alta	8,5%
8	Medicina Nuclear	Rosa	Alta	3,9%

Fonte: Elaboração própria.

2.3.2 Processando os chamados

A central de atendimento opera de segunda à sábado, das 8:00 às 18:00. O sistema de filas utilizadas pela a central de marcação é o PEPS (Primeiro Entra, Primeiro Sai), entretanto ocorrem priorizações na fila, de acordo com as etiquetas empregadas anteriormente pela a URA. Tais priorizações são mostradas no Quadro 3, onde o maior valor de prioridade significa maior prioridade frente a um chamado de menor valor.

Quadro 3: Prioridades entre os chamados

Fila	Prioridade
Ressonância Magnética	10
Medicina Nuclear	10
Oncologia Clínica	8
Radioterapia	8
Tomografia	8
Exame	5
Consulta	3
Ambulatório Caicó	3

Fonte: Elaboração própria.

A justificativa para o critério de priorização é em face que a categoria de exames possui uma disponibilidade maior de agenda em um período recente, visto que o hospital dispõe de um volume maior de maquinário e corpo de enfermeiros, frente as consultas que possuem um grau de ocupação superior. Além disto, existe uma priorização por chamadas de alta complexidade, que são encaminhadas somente para os operadores que tenham a proficiência em atendê-las.

Os chamados, após gerados e classificados, aguardam na fila para serem processados pelos agentes. Uma vez que o agente esteja disponível para o processamento, o chamado é processado. Foi averiguado com base nos dados históricos que o tempo de processamento varia conforme o tipo do chamado. Logo, para refletir esta realidade no modelo, cada tipo de chamado teve uma curva de distribuição própria atribuída a ele, semelhante ao processo realizado quanto aos tempos de chegada, conforme apresentado no Quadro 4.

Quadro 4. Distribuição de tempo de processamento dos chamados.

Tipo de Chamado	Distribuição de probabilidade
Consulta	Gamma[0, 115,382655, 1,492644]
Exame	Gamma[0,647343, 171,963551, 1,185725]
Oncologia Clínica	Weibull[0, 125,212259, 1,284235]
Radioterapia	Beta[0,995648, 788,283406, 0,921933, 8,761499]
Ambulatório Caicó	Pearsont6[2,956832, 1078,114455, 1,483433, 9,391463]
Tomografia	Beta[0,000949, 1339,440241, 1,012816, 6,604057]

Ressonância Magnética	Gamma[1,9637379, 90,730909, 1,422146]
Medicina Nuclear	Johnsonbounded[0,020080, 848,350743, 1,891728, 1,081919]

Fonte:Flexsim

2.3.3 Quadro de Funcionários e Pausas

O setor de marcação conta com vinte e dois operadores distribuídos entre quatro escalas diferentes, além de dispor de uma supervisora e uma diretora de operações. O setor possui um total de dezessete estações de trabalho. Quanto a designação de função, 12 funcionários são designados para marcações de baixa complexidade, 9 são designados para marcação de alta complexidade, destes, seis são polivalentes e realizam marcações de baixa complexidade quando não há demanda para alta complexidade, e por fim um único funcionário é designado para atividades de reagendamento.

Como o escopo deste estudo engloba os chamados de origem por meio do telefone, desconsiderou-se os operadores que trabalham unicamente com o as telemensagens, assim como o operador responsável pelo reagendamento, ficando a disponibilidade de agentes como mostrado no Quadro 5.

Quadro 5. Escala de horários e grupos de operadores.

Horários	De	Até	Grupo			Total
			A	B	C	
Tarde 1	12:00	18:30	0	0	2	2
Tarde 2	13:00	19:20	1	2	4	7
Manhã	07:00	13:20	2	1	3	6
Diurno	07:00	16:45	1	0	0	1
Intermédio	10:00	16:20	0	1	1	2

Fonte: Elaboração própria.

Afim de simplificar as competências para atender os diversos chamados, foram elaborados três grupos de operadores conforme suas habilidades. O grupo “C” de operadores atende chamadas classificadas como de Baixa Complexidade. O grupo “A” atende chamadas de Alta Complexidade, e o grupo “B” atende os dois tipos de chamada.

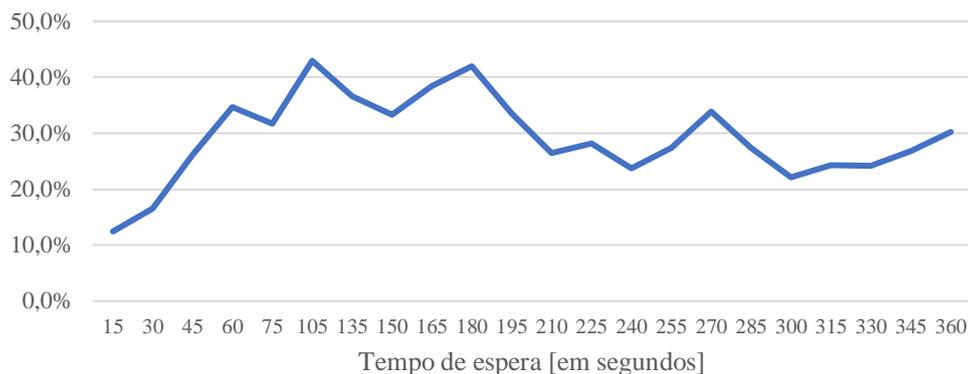
Estes grupos de funcionários foram implementados nos modelos assim como seus horários de trabalho. As pausas adotadas seguiram a NR 17 para operadores de telemarketing, que prevê que as pausas devem ser concedidas: a) fora do posto de trabalho; b) em 02 (dois) períodos de 10 (dez) minutos contínuos; c) após os primeiros e antes dos últimos 60 (sessenta) minutos de trabalho em atividade de teleatendimento/telemarketing. Além disso, é importante notar que o operador tem direito, de acordo com §1º do Artigo 71 da CLT, a um intervalo de 20 minutos para refeição.

Diante disto, para título de simplificação, optamos por modelar no sistema duas pausas de 20 minutos cada, tendo a primeira início uma hora após o início a atividade, e a segunda uma hora antes da finalização das atividades.

2.3.4 Abandono

Existe uma distinção entre o tempo que um cliente precisa esperar na fila e o tempo que ele está disposto a esperar antes de abandonar o sistema. Ambas as medidas são de grande importância, mas nenhuma é diretamente observável e, portanto, devem ser estimadas [Brown et al, 2005]. Neste estudo optamos por utilizar um critério por base nas observações dos abandonos prévios dos clientes. Ao inferir que os clientes abandonam o sistema devido ao tempo elevado de fila, foi analisada a porcentagem de abandonos em relação ao tempo de espera, conforme mostrado na Figura 5.

Figura 5. Chance de Abandono x Tempo de Espera.



Embora cada cliente seja distinto e, portanto, desafiador de se medir uma “paciência” teórica. Usou-se como convenção esta porcentagem de chance, que cresce a partir do primeiro segundo de espera e atinge seu pico máximo aos 105 segundos, de 43%, esta chance, então foi utilizada como convenção para os períodos consequentes, até uma espera máxima de 30 minutos, de onde o cliente abandonará a fila com certeza, pois o sistema realiza uma operação de timeout automática.

2.4 Validade do modelo e quantidade de replicações

Definido o modelo, buscou-se um nível de confiança de 95%, que infere em um nível de significância de 0,05. Para avaliar se os resultados do modelo de simulação se enquadram dentro do intervalo de confiança desejado, utilizou a metodologia proposta por Law e Kelton (2014), que apresenta a Equação 1 demonstrada a seguir, tomando como a variável de validação o tempo de espera da fila de baixa prioridade, visto que se trata de uma medida que flutua em consequência das escolhas adotadas na modelação do sistema.

$$\text{Erro Padrão} = t_{1-\frac{\alpha}{2}, n-1} * s / \sqrt{n} \quad (\text{Equação 1})$$

Donde:

t = Distribuição t para $1-\alpha$ com n-1 graus de liberdade.

s = Desvio padrão das médias das replicações do modelo.

n = Número de replicações/ Número de amostras.

Para título de testes, foi realizado uma rodada com 10 replicações, onde cada uma simulava 60 dias de trabalho. A média do tempo de espera para a fila de baixa prioridade neste intervalo foi de 187,89 segundos, onde o menor e o maior valor foram respectivamente 179,50 e 196,95 segundos. O desvio padrão da amostra foi de 6,86 segundos. Diante disto, realizado o cálculo proposto na Equação 1, considerando que o grau de liberdade é 9, e o *t de student* tem valor de 2,262, encontra-se uma precisão da amostra piloto de 5,17 segundos. Que por sua vez, é inferior a precisão desejada de 9,39 segundos (5% da média amostral), o que localiza as variações da simulação em um intervalo adequado de incerteza.

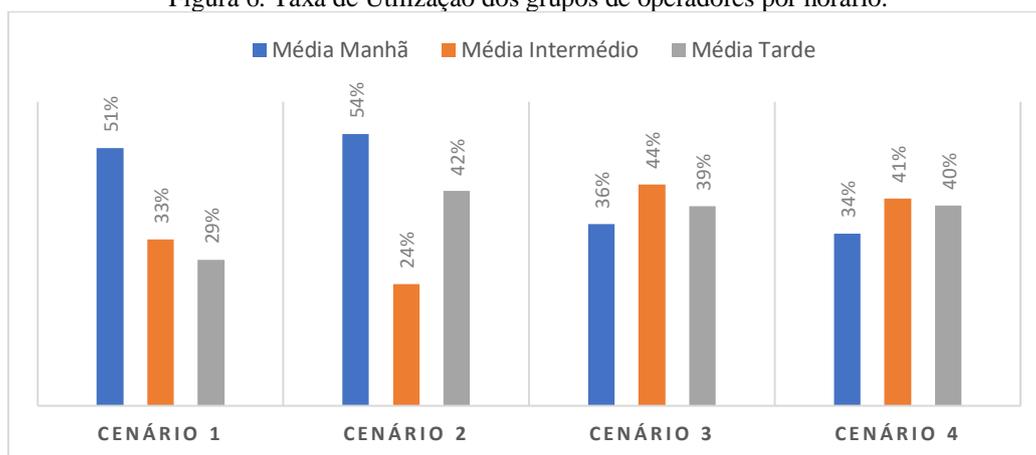
3 RESULTADOS

Foram elaborados quatro cenários que alteravam os seguintes aspectos da simulação: O cenário 1 corresponde a situação real, sem nenhuma alteração. O cenário 2 implementa o *crosstraining*, isto é, considera o treinamento dos operadores pertencentes ao grupo “A” para realizar atividades de baixa complexidade, desta forma transformando-os em operadores do grupo “B”. O cenário 3, implementa o *crosstraining* apresentado no cenário 2, e além disto, realoca três trabalhadores pertencentes ao grupo C e um pertencente ao grupo B que operam pela tarde para a manhã. O cenário 4, implementa as medidas supracitadas e gera uma nova regra de priorização, com base no menor *makespan*. Ou seja, as chamadas que são historicamente mais rápidas serão atendidas primeiro.

Os indicadores de performance utilizados na análise serão o tempo médio de espera para o atendimento e a porcentagem de abandono, além disto, este trabalho também fez uso da taxa de utilização dos diferentes grupos de operadores para avaliar a distribuição do volume de trabalho, e identificar possíveis gargalos na operação.

A Figura 6 apresenta a análise pela a dimensão dos turnos do trabalho. No cenário atual é possível perceber que os trabalhadores que operam pela manhã possuem uma carga de trabalho muito superior aos que operam no turno da tarde. Os cenários 3 e 4 são capazes de mitigar este desbalanceamento de carga de trabalho.

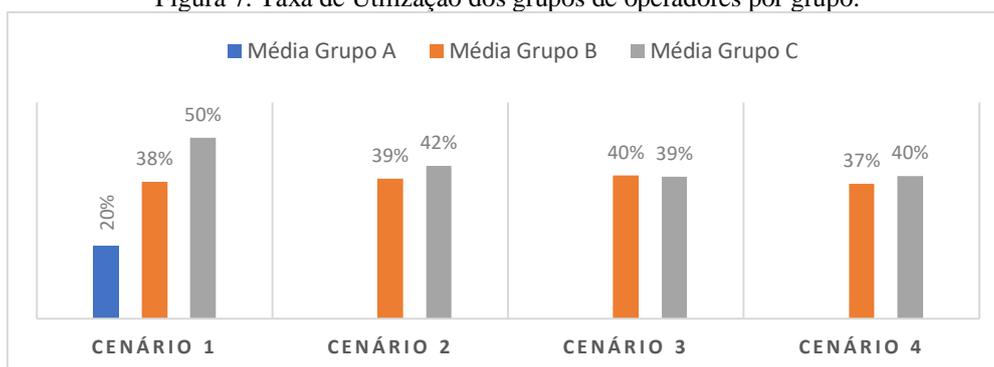
Figura 6. Taxa de Utilização dos grupos de operadores por horário.



Fonte: Elaboração própria.

Na análise sobre a ótica dos grupos de trabalho, (Figura 7) o cenário atual demonstra que os funcionários pertencentes ao grupo C possuem uma taxa de utilização até duas vezes maior que os pertencentes ao grupo A. A partir do cenário 2, o grupo A é extinto, visto que o *crosstraining* habilita este grupo a atender chamados de baixa complexidade, migrando estes funcionários ao grupo B. No cenário 2 é possível perceber que os dois grupos apresentam uma taxa de utilização equilibrada.

Figura 7. Taxa de Utilização dos grupos de operadores por grupo.



Fonte: Elaboração própria.

Referente aos tempos de espera, as mudanças implementadas ao longo dos cenários otimizam o critério do tempo de espera, tanto para as filas de baixa complexidade como para as filas de alta complexidade, como pode ser observado no quadro 6.

.Quadro 6: Resumo de Resultados.

Cenário	1 [Atual]	2	3	4
---------	--------------	---	---	---

Tempo médio de espera em fila de alta prioridade [em segundos]	23	28,7	24,9	39,5
Tempo médio de espera em fila de baixa prioridade [em segundos]	187,6	99,9	58,3	55
Taxa de Abandono	20%	17%	15%	15%

Fonte: Elaboração própria.

No quadro 6, observa-se que o cenário 2- que prevê o *crosstraining*- possibilita a redução no tempo de espera dos chamados de baixa complexidade em **47%**, o que é lógico, visto que aumenta a oferta de operadores capacitados a atender este tipo de chamado. Entretanto, ao ocupar os operadores mistos com tarefas de baixa complexidade, aumenta-se o tempo de espera da alta complexidade. O cenário 3 adota as medidas do cenário 2 e altera os turnos de quatro colaboradores para a manhã, vista a demanda maior neste período. Estas alterações são capazes de garantir uma redução do tempo de espera de **69%** nos chamados de baixa complexidade.

O cenário 4, que adota todas as alterações anteriores e acrescenta uma nova regra de priorização baseada no menor *makespan*, indica uma redução de **71%** do tempo de espera dos chamados de baixa complexidade, a maior redução entre os cenários, embora cause o aumento no tempo de espera em **72%** dos chamados de alta complexidade, isto se dá pela a perda de prioridade deste chamados garantido pelo o modelo de priorização anteriormente adotado.

Como no modelo de simulação, a taxa de abandono está ligada ao tempo de espera, implica que a diminuição dos tempos de espera acarreta na diminuição da taxa de abandono. No modelo adotado, baseado nos dados históricos, a taxa mínima de abandono registrada é a de 12,5%, ou seja, mesmo no tempo de espera zero, clientes ainda abandonam. Isto se dá a fatores externos aqueles da performance do sistema, como erros ao selecionar a opção correta na URA, erros ou atrasos na transmissão da chamada para o operador, telefone mudo, entre outros.

Feitas as considerações, é possível observar que as mudanças realizadas no cenário 2 possibilitaram a queda de 3% nos abandonos médios em relação ao cenário atual, as medidas implementadas no cenário 3 e 4 reduzem os abandonos em 5%.

Conforme todas as análises realizadas, o cenário 3 mostra-se o mais estrategicamente viável para a organização, pois diminui o tempo de espera significativamente para a maioria dos chamados e distribui mais equitativamente a taxa de utilização dos operadores.

4 CONCLUSÃO

O presente trabalho cumpre seu objetivo inicial de modelar uma central de teleatendimento médico por meio de uma simulação e gerar uma realocação de agentes que buscasse a diminuição do tempo de espera dos pacientes assim como a taxa de abandono de chamados. As medidas supracitadas, conforme discutidas com a gerência, são possíveis de serem adotadas dentro da realidade da organização e estão em processo de implementação. Os benefícios esperados indicam um balanceamento mais efetivo da carga de trabalho, a redução de até 69% do tempo de espera para chamados de baixa complexidade e uma redução de 5% dos abandonos.

Destaca-se como limitação deste estudo de simulação algumas simplificações feitas no modelo, como o fato de todos os agentes pertencentes a uma escala de horário tirarem suas pausas simultaneamente, fato que não ocorre na realidade da empresa, onde os operadores revezam suas pausas para manter o funcionamento do setor em toda a escala de trabalho. Além disto, por ausência de uma planta baixa atualizada das instalações, as acomodações foram feitas por meio de estimativas, de forma que o tempo de locomoção até a sala de repouso possa ter impactado a verossimilhança da simulação. Outro fator a se destacar é a não inclusão dos atendimentos via mensagens eletrônicas que também é realizado no setor, por motivo da plataforma de gerenciamento destes atendimentos não fornecer um relatório de dados consistentes. Destaca-se como sugestão de trabalhos futuros o acompanhamento das medidas adotadas no setor de teleatendimento e a comparação das performances esperadas com os resultados reais obtidos.

Referências

BOUZADA, Marco Aurélio Carino. Scenario Analysis within a Call Center Using Simulation. *Journal of Operations and Supply Chain Management*, v. 2, n. 1, p. 89, 2009.

Stier

BREUER, Lothar; BAUM, Dieter. An introduction to queueing theory: and matrix-analytic methods. Springer Science & Business Media, 2005.

BROWN, Lawrence et al. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, v. 100, n. 469, p. 36–50, 2005.

ERDEM, Aslı Sencer; ALAÇAM, Serkan. SIMULATION MODELING FOR CALL CENTER MANAGEMENT: A CASE STUDY IN A PRIVATE BANK, 2007.

FOGLIATTI, M.C.; MATTOS, N.M.C. Teoria de Filas. Rio de Janeiro: Interciência, 2007.

FROST, Y. SULLIVAN. Latin America Contact Center Outsourcing Services Markets. 2008.

HILLIER, Frederick S.; LIEBERMAN, Gerald J. Introdução à pesquisa operacional. McGraw Hill Brasil, 2013.

INSTITUTO NACIONAL DO CÂNCER. Estimate/2018 – Cancer Incidence in Brazil. [S.l: s.n.], 2017.

KLEINROCK L. Queueing Systems. Vol. 1: Theory. New York: John Wiley & Sons, Inc. 1975.

KOOLE, Ger; MANDELBAUM, Avishai. Queueing models of call centers: An introduction. *Annals of Operations Research*, v. 113, n. 1–4, p. 41–59, 2002.

LAMARCA, Braz Rafael da Costa. Utilização da simulação para avaliação de equipes cirúrgicas de revascularização miocárdica em relação ao Australian National Diagnosis Related Group [AN-DRG] em um hospital privado. Dissertação [Mestrado em Engenharia de Produção]. Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2008.

LAW, A. M.; KELTON, W. D. SIMULATION MODELING & ANALYSIS. 5. ed. McGraw-Hill, 2014.

MEHROTRA, Vijay; FAMA, Jason. Call center simulation modeling: Methods, challenges, and opportunities. *Winter Simulation Conference Proceedings*, v. 1, n. May, p. 135–143, 2003.

OWENS, Katie; STIER, Richard. The intersection of contact centers and patient experience. *ANSWERStat: Healthcare Contact Center Information Hub*. 27 de julho de 2016.

RIO GRANDE DO NORTE. Secretaria de Estado da Saúde Pública. Coordenadoria de Planejamento e Controle de Serviços de Saúde. Plano Estadual de Saúde – PES: 2016 -2019. Rio Grande do Norte, 2016.

STIER, Richard D. The medical call center: Marketing health services, v. 19, n. 2, p. 25, 1999.