



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**Análise comparativa de mecanismos de priorização de
atendimento médico por teoria das filas: Um estudo de
caso no setor de emergência do Hospital da Restauração
de Pernambuco**

André Luís de Oliveira Cardoso Coelho

Orientador: Prof. Rodrigo Ferreira

Recife, Dezembro/2010

André Luís de Oliveira Cardoso Coelho

Análise comparativa de mecanismos de priorização de atendimento médico por teoria das filas: Um estudo de caso no setor de emergência do Hospital da Restauração de Pernambuco

Trabalho de Conclusão de Curso apresentado na Universidade Federal de Pernambuco (UFPE) como requisito parcial para obtenção do grau de bacharel em Engenharia de Produção.

Orientador: Prof. Rodrigo Ferreira

Recife, Dezembro/2010

C672a Coelho, André Luís de Oliveira Cardoso.

Análise comparativa de mecanismos de priorização de atendimento médico por teoria das filas: um estudo de caso no setor de emergência do Hospital da Restauração de Pernambuco / André Luís de Oliveira Cardoso Coelho. - Recife: O Autor, 2010.
viii, 52 folhas; il., tabs.

TCC (Graduação) – Universidade Federal de Pernambuco.
CTG. Curso de Engenharia de Produção, 2010.
Orientador: Prof. Rodrigo Ferreira.
Inclui Referência.

1. Engenharia de Produção. 2. Teoria das Filas. 3. Modelos de Priorização do Atendimento Médico. I. Ferreira, Rodrigo.
II. Título

658.5 CDD (22. ed.)

UFPE/BCTG/2010-248

AGRADECIMENTOS

Agradeço a meus pais, Giliate e Cristina, que nunca mediram esforços para minha educação. Desde pequeno, sempre me mostrando e acreditando o quão é importante o estudo na vida de uma pessoa.

Agradeço a toda minha família, meus irmãos, minha irmã, avós, avôs (in memoriam), tios e tias, pela paciência, atenção e orientações ao longo dessa minha jornada acadêmica. Em especial ao meu avô Paulo, que apesar de não estar mais fisicamente entre nós, sempre senti sua presença ao meu lado.

Ao meu orientador, Rodrigo Ferreira, que entre idas e vindas entre Natal e Recife, sempre esteve disposto a me ajudar com este trabalho, me guiando e orientado da melhor forma possível.

Aos demais Professores e Professoras do Departamento de Engenharia de Produção da UFPE e também à própria UFPE por me acolher no meio acadêmico.

A todos os meus amigos e amigas que fiz durante o curso de Engenharia de Produção e que esta amizade possa vencer os limites geográficos da Universidade.

À Dra Fátima Buarque que me guiou no processo de obtenção dos dados deste trabalho.

À Ana Cláudia do Comitê de Ética em Pesquisa do Hospital da Restauração, pela especial atenção e dedicação que me foram dadas. E a todo o Comitê também.

Às Recepcionistas do Hospital da Restauração e em especial à Patrícia Barros, que entre suas tarefas corriqueiras do dia a dia, dedicou-me tempo e disposição para me suprir com os dados que fundamentaram este trabalho.

Ao meu amigo Raoli, que apesar de todas as dificuldades encontradas na vida, foi capaz de superá-las de uma forma que eu jamais conseguiria.

RESUMO

As longas esperas nas filas dos serviços de emergência dos hospitais representam o mau uso dos recursos disponíveis e um mau dimensionamento da capacidade do sistema. Buscando oferecer um serviço mais humanitário, diversos hospitais adotaram modelos de priorização do atendimento. Esses modelos foram desenvolvidos no intuito de classificar os pacientes em categorias e priorizar o atendimento daqueles que mais necessitam de rapidez para iniciarem seu atendimento médico, a fim de não agravarem seu estado clínico. Entretanto, os demais pacientes classificados dentro das categorias de menor risco de vida, teriam que aguardar mais tempo para iniciar o atendimento. Este trabalho levou em consideração dois fatores: os modelos mais utilizados de priorização do atendimento médico no serviço de emergência dos hospitais públicos e a Teoria das Filas. Como conclusão percebeu-se que o modelo de priorização de Manchester melhor se adapta às características e particularidades do Hospital da Restauração de Pernambuco, estudo de caso deste trabalho. Além disso, resultados significativos foram encontrados a partir da modelagem feita pela teoria das filas no que envolve os parâmetros da fila do Hospital da Restauração.

Palavras chave: Teoria das filas, Modelos de priorização do atendimento médico, serviço de emergência em saúde.

SUMÁRIO

1. Introdução.....	1
1.1. Objetivos.....	3
1.1.1. Objetivo geral.....	3
1.1.2. Objetivos específicos.....	4
1.2. Estrutura do trabalho.....	4
1.3. Metodologia.....	4
2. Modelos de priorização em atendimentos do serviço de emergência em saúde.....	6
2.1 Modelo de Priorização de Manchester.....	9
2.2 Modelo de Priorização Australiano.....	15
2.3 Modelo de Priorização Canadense.....	17
2.4 Modelo de Priorização Americano.....	20
2.5 Comparação entre os modelos de Priorização.....	22
3. Teoria das filas.....	24
3.1 Estrutura básica dos modelos de filas.....	24
3.1.1 Processo de filas básico.....	24
3.1.2 Fonte de entradas (população solicitante).....	25
3.1.3 Tempo de serviço.....	26
3.1.4 Fila.....	26
3.1.5 Disciplina das filas.....	26
3.1.6 Mecanismo de atendimento ou canais de serviço.....	27
3.1.7 Processos de filas elementar.....	27
3.1.8 Notação.....	28
3.1.8.1 Relações entre L , L_q , W e W_q e fórmula de Little.....	30
3.1.9 Medidas de desempenho.....	31
3.2 Modelos de filas baseados no processo de nascimento-e-morte.....	31
3.2.1 Modelo M/M/s.....	32
3.2.1.1 Caso com um único atendente ($s = 1$) (M/M/1).....	33
3.2.1.2 Caso com vários atendentes ($s > 1$) (M/M/s).....	34
3.2.2 Modelos de filas de disciplinas de prioridades.....	36
3.2.2.1 Com prioridades não-preemptivas.....	37

3.2.2.2 Com prioridades preemptivas.....	37
4. Estudo de caso.....	39
4.1 O Hospital da Restauração de Pernambuco.....	39
4.2 As Unidades de Pronto-Atendimento (UPAs).....	39
4.3 Coleta e análise dos dados.....	40
4.4 Modelo e filas no HR.....	44
4.5 Análise de resultados.....	49
5. Conclusão.....	50

BIBLIOGRAFIA

ANEXO A

LISTA DE FIGURAS

Figura 1.1	Comparação entre o serviço de urgência/emergência e um sistema de produção.....	3
Figura 2.1	Classificação das categorias do MTS.....	11
Figura 2.2	Fluxograma orientador para a queixa principal tipo “dor cervical”, segundo o MTS.....	15
Figura 2.3	Algoritmo disposição do fluxograma do ESI.....	21
Figura 3.1	Representação básica de um sistema de filas.....	25
Figura 3.2	Representação do processo de filas elementar.....	28
Figura 3.3	Diagrama de taxas para o modelo M/M/s ($s = 1$).....	32
Figura 3.4	Diagrama de taxas para o modelo M/M/s ($s > 1$).....	33
Figura 4.1	Total de pacientes atendidos durante o período de 12/11/2010 à 18/11/2010 classificados de acordo com sua gravidade e cor do Sistema de Triagem de Manchester.....	42
Figura 4.2	Total de pacientes na semana x hora do dia.....	43
Figura 4.3	Média da taxa de chegada/hora nos dias em horário de pico.....	44

LISTA DE TABELAS

Tabela 2.1	Índice do fluxograma de decisão do MTS	14
Tabela 2.2	Relação entre nível, descrição categoria e tempo de atendimento do ATS.....	17
Tabela 2.3	Relação entre nível, características e tempo de atendimento do CTAS.....	19
Tabela 2.4	Comparações entre os modelos ATS, CTAS, MTS e ESI.....	22
Tabela 2.5	Comparações entre os tempos máximos previstos em cada categoria do modelos MTS, CTAS E ATS.....	23
Tabela 4.1	Quantidade de pacientes classificados e atendidos por dia.....	41
Tabela 4.2	Porcentagem do total de pacientes e taxa de chegada nas cinco categorias de Manchester.....	45
Tabela 4.3	Resultados do modelo de disciplina de prioridades para $s = 1$	46
Tabela 4.4	Resultados do modelo de disciplina de prioridades para $s = 2$	47
Tabela 4.5	Resultados do modelo de disciplina de prioridades para $s = 3$	48

1. INTRODUÇÃO

Com mais de 190 milhões de brasileiros em território nacional, os hospitais, as clínicas, pronto socorros, as unidades de atendimento de urgência e emergência em geral, não conseguem suprir a demanda que envolve a quantidade de atendimento de pacientes no serviço de saúde pública, além do que, na maioria dos casos, o serviço do atendimento é inadequado, há longas esperas nas filas, fluxo desordenado de pessoas e percebe-se um mau uso dos recursos disponíveis, caracterizando como um grave problema na sociedade.

Criado em 1988 pela Constituição Federal Brasileira, o Sistema Único de Saúde (SUS) é um dos maiores sistemas de saúde pública do mundo. Com garantia de acesso integral, universal e igualitário para toda população brasileira, além de descentralizado (a gestão do sistema de saúde passa para os municípios, com a consequente transferência de recursos financeiros pela União, além da cooperação técnica). Sua principal e preferencial porta de entrada são os postos de saúde, centros de saúde e unidades de Saúde da Família, onde, a partir do primeiro atendimento, o paciente é encaminhado e direcionado para hospitais e clínicas especializadas, responsáveis por serviços de maior complexidade (BRASIL, 2007).

Os Municípios, porém, não têm condições de ofertar na integralidade os serviços de saúde, portanto, para que o sistema de atendimento funcione é necessária uma estratégia regional de atendimento, traduzida numa parceria entre o Estado e os Municípios. Essa estratégia é caracterizada pela referencialização (conceito básico do SUS) do atendimento. Para cada tipo de enfermidade há um local de referência para o atendimento. Um segundo conceito básico do SUS é a hierarquização da rede, ou seja, a ideia é que haja centros de referência para graus de complexidade diferentes de serviços. Caso um paciente necessite passar por uma cirurgia e seu município não possua atendimento hospitalar, será encaminhado para um hospital de referência em uma cidade vizinha, de acordo com o plano regional de serviços, traçado pelos gestores municipais e estaduais (BRASIL, 2007).

No que se refere ao SUS, não existem avaliações sistemáticas sobre prazos de espera para internações, consultas, exames ou o número de pessoas que esperam pelo atendimento nas filas, nem tampouco avaliações oficiais dos custos desses procedimentos. Realidade bem diferente do que ocorre em países mais desenvolvidos como Canadá, Inglaterra, Estados Unidos e Austrália, onde há inúmeras publicações de dados detalhados a respeito deste tema. Resultado do descompasso entre demanda e oferta, as filas no SUS são originadas devido a três causas (MARINHO, 2009):

- Governamental: responsável pelo tamanho do orçamento geral da saúde
- Autoridades Individuais e Instituições médicas, científicas, jurídicas e empresarias atuantes no setor: responsáveis por decidir sobre os benefícios e os custos das internações
- Profissionais de Saúde, principalmente os médicos: responsáveis por decidir quais são as necessidades clínicas do paciente.

As longas filas enfrentadas pelos pacientes no SUS causam uma espera prolongada no atendimento, conseqüentemente, geram-se impactos negativos no bem-estar desses pacientes, nas probabilidades de cura, na natureza e extensão das sequelas. E não somente ao paciente propriamente dito, mas também em seus familiares, além de gerar importantes custos adicionais ao sistema, obrigando-o a arcar com pesados ônus administrativos e perda da eficiência.

Diretamente relacionado com a problemática das filas no atendimento do serviço de saúde pública no Brasil, existe a questão da priorização no atendimento dos pacientes, ou seja, quando o paciente chega ao serviço de saúde, este é submetido a um processo de triagem e em seguida classificado de acordo com seu quadro clínico ou outros fatores. Esta classificação resultará num tempo previsto máximo para o paciente ser atendido.

A priorização na escolha de pacientes é uma das grandes dificuldades enfrentada pela equipe médica, em função da grande demanda e do número limitado de vagas. Alguns estudos já foram desenvolvidos nesse tema envolvendo modelagem de problemas com o apoio de diversos métodos de decisão (ALMEIDA & ALMEIDA, 2003).

O presente documento tem como propósito ampliar os conhecimentos disponíveis acerca dos temas abordados e fazer um estudo comparativo entre modelos de priorização de atendimento de pacientes, no intuito de melhorar a qualidade do atendimento prestado pelo SUS, evitando-se, desta forma, desperdícios de recursos quando estes poderiam ser melhor utilizados.

A escolha do tema deve-se a afinidade do autor com as contribuições da Engenharia de Produção para com a área de saúde (afinidade esta por se tratar de que os pais do autor são profissionais da área da saúde e, desta maneira, desejou-se com este trabalho, proporcionar uma contribuição da Engenharia de Produção à área de saúde) além da inquietação constante em saber que pessoas morrem todos os dias no mundo em filas à espera de um atendimento nos serviços públicos de saúde.

Analizando o serviço de urgência e emergência de um hospital, pode-se compará-lo como um sistema de produção. Por exemplo, num sistema de produção tem-se uma matéria prima (insumo) sofrendo um processamento e transformando-se num produto final (saída). De acordo com esta análise, num serviço de urgência e emergência de um hospital, têm-se os insumos (caracterizados pelos pacientes, informações, energia, etc.), a seguir o processamento destes insumos, caracterizado pelo atendimento médico e, por final, sua saída (caracterizada pelo cliente devidamente atendido).

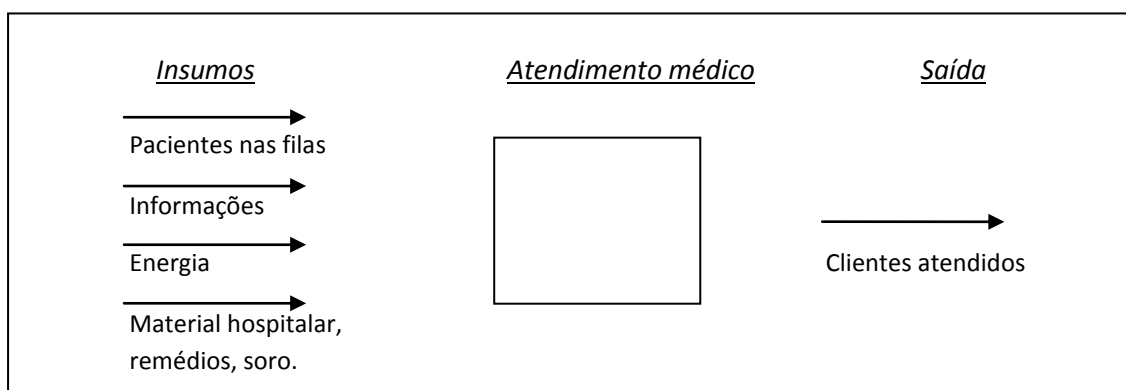


Figura 1.1 – Comparação entre o serviço de urgência/emergência e um sistema de produção

Fonte: O autor (2010)

1.1. Objetivos

1.1.1. Objetivo Geral

Desenvolver um estudo que auxilie na melhoria da qualidade do atendimento de sistemas de emergência em saúde pública através da proposição de mecanismos de priorização e dimensionamento da capacidade do sistema.

1.1.2. Objetivos Específicos

1. Buscar na literatura modelos de priorização de atendimento a clientes no sistema de saúde pública
2. Comparar Modelos de priorização de atendimento de pacientes
3. Aplicação de um estudo de caso

1.2. Metodologia

Este trabalho foi um estudo de caso, desenvolvido no Hospital da Restauração de Pernambuco e devidamente autorizado pelo Comitê de Ética em Pesquisa do Hospital da Restauração (CEP/HR) através do nº CAAE - 0117.0.102.000-10 (Anexo A). Não foi realizado nenhum tipo de questionário ou entrevista.

Realizou-se uma análise gráfica de dados, em conjunto com conceitos e modelos básicos de filas para análise e avaliação.

Foram utilizados modelos clássicos da Teoria das Filas para avaliação dos vários parâmetros da fila. Realizou-se uma coleta de dados estatísticos referentes à taxa de chegada dos pacientes por hora e por cada categoria do Sistema de Triagem de Manchester na emergência do Hospital da Restauração, assim como o tempo, aproximado, do atendimento médico clínico. Os dados coletados foram referentes ao período de 12 a 18 de novembro do ano de 2010. Tais dados foram fornecidos somente após a autorização do CEP/HR e entregues em forma de relatórios impressos pelo Sistema de Informação do hospital ao autor.

1.3. Estrutura do Trabalho

O presente trabalho foi estruturado da seguinte maneira:

- Capítulo 1 – Introdução. Foi apresentada neste capítulo uma introdução sobre os temas abordados deste trabalho, assim também como os objetivos deste estudo, a estrutura do trabalho e a metodologia.
- Capítulo 2 – Modelos de priorização em atendimentos do serviço de emergência em saúde. Neste capítulo foram apresentados alguns aspectos gerais sobre os principais modelos de priorização e as características e particularidades desses modelos, assim também como suas diferenças.
- Capítulo 3 – Teoria das filas. Neste capítulo foram apresentados conceitos sobre Teoria das filas, os elementos que compõem um modelo de filas básico e alguns modelos desta teoria.
- Capítulo 4 – Estudo de caso. Apresentaram-se neste capítulo algumas considerações sobre o Hospital da Restauração de Pernambuco, estudo de caso deste trabalho, dados pertinentes à coleta e, resultados da aplicação do modelo da Teoria das Filas na fila do hospital.
- Capítulo 5 – Conclusão. Este capítulo apresentou as conclusões que o autor chegou através da análise dos resultados e, assim também como a contribuição deste trabalho para o autor.

2. MODELOS DE PRIORIZAÇÃO EM ATENDIMENTOS DO SERVIÇO DE EMERGÊNCIA EM SAÚDE

Segundo Rati (2009) a procura pelo atendimento de urgência e emergência por pacientes que, do ponto de vista técnico, não necessitam, necessariamente, desse tipo de atendimento tem sido descrito na literatura nos últimos anos. A atitude dos usuários ao buscarem hospitais, pronto socorros e pronto atendimento, como a primeira opção para atendimentos não considerados de urgência ou emergência, vem sobrecarregando esses serviços e preocupando profissionais de saúde e gestores do sistema.

Modelos de priorização em atendimentos de urgência e emergência do serviço de saúde pública são vistos como ferramentas potenciais para reduzir as superlotações nos hospitais e proporcionar um serviço de maior qualidade. Porém, não é intenção neste trabalho aprofundar o conhecimento teórico sobre tais modelos, contudo, a fim de que os objetivos deste estudo sejam alcançados, o autor considera de significativa importância uma breve introdução clara e objetiva sobre a funcionalidade e características particulares dos modelos mais utilizados nos serviços de urgência e emergência dos principais hospitais públicos nacionais e internacionais.

Os modelos de priorização em atendimentos podem ser encontrados também na literatura, muito comumente, como triagem de priorização. Segundo Silva (2009) a palavra triagem tem sua origem do verbo *trier* que significa escolher, separar, selecionar. No processo de triagem de prioridades são exigidos alguns requisitos, tais como: capacidade de interpretação, discriminação e avaliação, raciocínio clínico, reconhecimento de padrões, formulação de hipóteses, representação mental e, menos importante, a intuição do tomador de decisão. Já no que se refere ao processo de tomada de decisão, três fases devem ser compreendidas:

1. Identificação de um problema
2. Determinação das alternativas
3. Seleção da alternativa mais adequada.

Desta forma, segundo o mesmo autor e considerando-se os descritos acima, diversos modelos de priorização de atendimento utilizam como base algoritmos de decisão como

opções estruturadas, facilitadoras do raciocínio e promotoras da uniformidade de critérios.

Segundo Moreira (2010), o processo de priorização do atendimento de pacientes não é um ponto final, mas o início de um processo de exame e discriminação clínica. A análise da implementação dos sistemas de triagem e de outros dados inerentes à mesma são de extrema relevância para o desenho e mudança operacional dos serviços de urgência e emergência das unidades de pronto atendimento.

Ainda de acordo com Moreira (2010), a avaliação médica de urgência/emergência e o tratamento inicial deverão estabelecer-se de acordo com um plano bem definido, o qual deverá ter por base as necessidades da comunidade e as capacidades reais dos recursos disponíveis de cada hospital. É neste contexto, que o procedimento da triagem surge, sendo crucial a sua implementação para garantir:

- ✓ Uma boa gestão de cada serviço de emergência/ urgência
- ✓ Equidade dos cuidados de saúde a prestar.
- ✓ Um melhor uso dos recursos disponíveis
- ✓ Um fluxo mais ordenado de pessoas dentro do serviço público de saúde
- ✓ Redução de custos
- ✓ Otimização das atividades prestadas
- ✓ Apoio no processo de tomada de decisão
- ✓ Criação de indicadores de qualidade
- ✓ Satisfação das necessidades do cliente

A utilização de modelos de priorização em todas as áreas de atenção às urgências tem inúmeras vantagens, conforme destaca Esteves (2009):

- O protocolo único de classificação de risco é reproduzível, passível de auditoria e com controle médico e promove o atendimento médico em função do critério clínico e não do administrativo ou da simples ordem de chegada ao serviço de urgência.
- A utilização em vários centros permite uma caracterização melhor da demanda de procura nas portas de urgência em todo o estado.

- Estabelecer um tempo de espera pela atenção médica e não um diagnóstico.
- O fluxograma da decisão parte da identificação da queixa inicial para, por fim, estabelecer o tempo de espera de acordo com a gravidade da condição do paciente.
- O sistema depende mais da disciplina em sua utilização do que da formação técnica de quem o utiliza.
- O sistema prevê a classificação individual de prioridades e garante a uniformidade de critérios ao longo do tempo entre as diversas equipes.
- O sistema é de rápida execução e permite eliminar os riscos de uma triagem feita por pessoas despreparadas (ex: porteiro, seguranças, etc.).
- A adoção do sistema não implica investimento financeiro significativo.
- Determina e orienta a reconfiguração do fluxo e da arquitetura nos serviços de urgência.

Para Iserson & Moskop (2007 apud Moreira, 2010), a utilização de modelos de priorização de atendimento de pacientes requer que três condições sejam satisfeitas:

- 1) Os recursos podem variar consideravelmente, desde o moderado em que nem todos os doentes podem ser atendidos de imediato, até ao momento após uma catástrofe em que centenas ou milhares de pessoas sofrem graves lesões num curto espaço de tempo. Em circunstâncias onde recursos seriam suficientes para servir todas as necessidades sem implicar, necessariamente, em espera no atendimento, a aplicação de modelos de priorização não seriam necessários. O outro extremo seria se não existissem recursos disponíveis no momento, tornando-se crucial a aplicação de tais modelos.
- 2) Geralmente, é o profissional de saúde, enfermeiro ou médico, que decide qual a necessidade clínica do paciente, baseando-se num breve atendimento inicial. Esta avaliação distingue a prática de triagem, na qual a decisão é específica para cada indivíduo.

- 3) Os modelos de priorização utilizam como base, algoritmos ou uma série de critérios, para determinar uma prioridade específica ou um tratamento específico para cada paciente.

Neste trabalho, foram abordados quatro dos principais modelos de priorização utilizados nos principais hospitais, são eles: Modelo de Manchester (Manchester Triage System - MTS), Modelo Australiano (Australasian Triage Scale - ATS), Modelo Canadense (Canadian Triage Acuity Scale - CTAS) e Modelo Americano (Emergency Severity Index - ESI). Conforme Cordeiro Júnior (2009), os quatro modelos de priorização são exemplificados da seguinte maneira:

- Modelo de Manchester (*Manchester Triage System* - MTS) - Trabalha com algoritmos e determinantes, associados a tempos de espera simbolizados por cor. Está sistematizado em vários países da Europa. O mecanismo de entrada é uma queixa do paciente;
- Modelo Australiano (*Australasian Triage Scale* - ATS) - Foi o pioneiro e usa tempos de espera de acordo com gravidade;
- Modelo Canadense (*Canadian Triage Acuity Scale* - CTAS) - Muito semelhante ao modelo australiano, é muito mais complexo e está em uso em grande parte do sistema canadense. O mecanismo de entrada é uma situação pré-definida;
- Modelo Americano (*Emergency Severity Index* - ESI) - Trabalha com um único algoritmo que foca mais na necessidade de recursos para o atendimento. Não é usado em todo o país.

2.1 Modelo de Priorização de Manchester (Manchester Triage System – MTS)

Atual modelo de priorização empregado no serviço de urgência e emergência do Hospital da Restauração, localizado na cidade do Recife, é um dos mais usados na maioria dos hospitais no mundo. Foi desenvolvido em 1997, no Hospital de Manchester, Inglaterra (GOMES, 2008).

Etapa crucial na gestão, o modelo de priorização de Manchester, caracteriza-se nas situações em que a demanda ocasionada pelo número de pacientes à espera do

atendimento excede a capacidade de atendimento do sistema. Dispondo de um sistema de triagem de forma objetiva e sistematizada, seu objetivo é promover o atendimento médico em função de critério clínico, e não do administrativo ou através da ordem de chegada dos pacientes no serviço de urgência e emergência. Tal sistema permite a identificação da prioridade clínica e definição do tempo recomendado até o início da primeira observação médica (SILVA, 2009).

O Sistema de Triagem de Prioridades de Manchester permite a identificação da prioridade clínica e a definição do tempo máximo recomendado até a primeira consulta médica. O objetivo é fazer a triagem de prioridades, identificando critérios de gravidade, de uma forma objetiva e sistematizada, que indicam a prioridade clínica com que o doente deve ser atendido e o respectivo tempo máximo recomendado até o atendimento médico (MACKWAY-JONES & MARSDEN, 2006).

Conforme Baumann & Strout (2007 apud Moreira, 2010), a triagem de Manchester é o sistema mais utilizado na Inglaterra, sendo uma representação baseada em fluxogramas. A primeira etapa diz respeito à pessoa responsável por priorizar os pacientes, identificar sua queixa principal e a partir de cinquenta e dois fluxogramas pré-determinados, selecionar o mais adequado. Em seguida é desenvolvida uma entrevista estruturada e assinalada uma categoria que irá do nível um ao nível cinco, identificada por cinco cores.

Silva (2009, p.36) destaca a essência do modelo de acordo com o trecho a seguir:

“O método consiste em identificar a queixa inicial (de apresentação) e seguir o respectivo fluxograma de decisão (existem ao todo 52 que abrangem todas as situações previsíveis). O fluxograma contém várias questões a serem colocadas pela ordem apresentada (com a definição exata dos termos) que constituem os chamados “discriminadores”. Estes podem ser específicos para a situação em causa (por exemplo, oftalmológica) ou gerais: perigo de vida, dor, hemorragia, estado de consciência, temperatura e o fato de se tratar ou não de uma situação aguda.

Perante a identificação do discriminador relevante determina-se a prioridade clínica (com a respectiva cor de identificação)”.

Após a identificação do quadro clínico, classifica-se o paciente em cinco categorias, identificadas por um número, nome, cor e o tempo máximo para ser atendido:

Vermelho	(1) EMERGENTE → Atendimento Imediato
Laranja	(2) MUITO URGENTE → Tempo de atendimento Máx. = 10min
Amarelo	(3) URGENTE → Tempo de atendimento Máx. = 60min
Verde	(4) POUCO URGENTE → Tempo de Atendimento Máx. = 120min
Azul	(5) NÃO URGENTE → Tempo de Atendimento Máx. = 240min

Figura 2.1 – Classificação das categorias do MTS

Fonte: Adaptado de Silva (2009)

A figura 2.1 ilustra as cinco classificações do MTS. Pacientes classificados como Emergente, devem ser identificados por uma pulseira na cor vermelha e devem ser atendidos imediatamente. Pacientes classificados na categoria Muito urgente devem ser identificados na cor laranja e devem esperar até 10 minutos para serem atendidos. Pacientes classificados na categoria Urgente tem 60 minutos para serem atendidos e são identificados na cor amarela. Pacientes identificados na cor verde são classificados na categoria Pouco urgente e esperam em até 120 minutos para começar o atendimento. Enquanto que na categoria Não urgente, os pacientes são identificados através da cor azul e devem aguardar até 2 horas para o atendimento.

A forma como Silva (2009) destaca as vantagens deste modelo é destacada da seguinte maneira:

- 1) Garantia da uniformidade de critérios ao longo do tempo e com as diversas equipes de serviço.

- 2) Fim da triagem do porteiro, que encaminha o doente sem critério objetivo, e permite a decisão tomada cientificamente, pondo de lado argumentos rudimentares, como por exemplo, que entrou de pé ou na maca, etc...
- 3) Não exige uma diferenciação especialmente exigente, mas sim um bom técnico de saúde e disciplina, geralmente realizado pelo enfermeiro (a). O sistema é institucional, protege o paciente que realmente necessita de urgência e o técnico de saúde. Esta solução facilita a gestão dos recursos disponíveis na medida em que não é preciso deslocar médico altamente qualificado para a triagem, sujeito a possuir critérios diversos e sem cobertura institucional para as decisões individuais.
- 4) Prevê a triagem individual (de doentes caso a caso), bem como as situações de exceção (com múltiplos doentes).
- 5) Não implica um investimento financeiro significativo.
- 6) É muito rápido de executar (com médias cronometradas por meios informáticos em diversos hospitais de 60 a 90 segundos por triagem).
- 7) Permite comparar dados entre os diversos hospitais que utilizam este modelo..

Segundo Rati (2009), estudos na literatura sobre a aplicabilidade do modelo de priorização de Manchester ainda são escassos e difíceis de encontrar. Alguns estudos sobre sua aplicabilidade em emergências pediátricas concluíram que tal modelo tem validade moderada, pois, tende a errar para o lado da segurança, apresentando muito mais resultados com índices de urgências maiores do que seus casos, respectivamente, apresentam. Desta forma, nota-se um número elevado de casos com urgência, onde, a princípio não deveria ser, sobrecarregando o sistema.

Para Baumann & Strout (2007 apud Moreira, 2010), o tempo médio para ser realizada a triagem de pacientes é de 2 a 5 minutos. Porém, quando este modelo estiver priorizando crianças no atendimento, este tempo poderá ser mais demorado, ultrapassando os minutos previstos inicialmente, buscando uma decisão mais precisa, visando não sobrecarregar o sistema.

Característica diferencial deste modelo é o uso de sistemas de informação. Toda informação, desde a chegada do paciente até sua saída, é processada e armazenada num banco de dados para posterior verificação e até mesmo uma possível auditoria da

operacionalização do modelo. De acordo com Esteves (2009) a informatização deste modelo de priorização conduz a um grande número de vantagens, conforme mostrado a seguir:

- Facilita a utilização pelo profissional que realiza a triagem das diversas variáveis que compõem o protocolo, diminuindo o tempo de triagem.
- Confere maior agilidade, facilidade e segurança no processo contínuo de auditoria que o protocolo prevê que ocorra com a escolha aleatória de cerca de 200 casos em média por mês a serem auditados.
- Evita erros no seguimento dos fluxogramas e obriga à boa prática nos registros e consequentemente nas informações geradas.
- Permite a mensuração adequada dos tempos e resultados dos atendimentos, permitindo comparações entre unidades e possibilitando o estabelecimento de padrões desejáveis.
- É fundamental na construção de uma linguagem única e no estabelecimento de uma rede integrada de serviços de saúde.

Tabela 2.1 – Índice do fluxograma de decisão do MTS

Fonte: Souza, 2009

1. Agressão
2. Asma
3. Autoagressão
4. Bebê que chora
5. Catástrofe: avaliação primaria
6. Catástrofe: avaliação secundaria
7. Cefaleia
8. Comportamento estranho
9. Convulsões
10. Corpo estranho
11. Criança com dificuldade de locomoção
12. Criança irritável
13. Criança que não se sente bem
14. Diabetes
15. Diarreia
16. Dispneia
17. Dispneia na criança
18. Doença hematológica
19. Doença mental
20. Doenças sexualmente transmissíveis
21. Dor abdominal
22. Dor abdominal na criança
23. Dor cervical
24. Dor de garganta
25. Dor lombar
26. Dor testicular
27. Dor torácica
28. Embriaguês aparente
29. Erupções cutâneas
30. Estado de inconsciência
31. Exposição a químicos
32. Feridas
33. Grande traumatismo
34. Gravidez
35. Hemorragia gastrointestinal
36. Hemorragia vaginal
37. Indisposição no adulto
38. Infecções locais e abscessos
39. Lesão toraco-abdominal
40. Mordeduras e picada
41. Pais preocupados
42. Problemas estomatológicos
43. Problemas nasais
44. Problemas nos membros
45. Problemas oftalmológicos
46. Problemas nos ouvidos
47. Problemas urinários
48. Queda
49. Queimaduras profundas e superficiais
50. Sobredosagem ou envenenamento
51. TCE- trauma crânio-encefálico
52. Vômitos

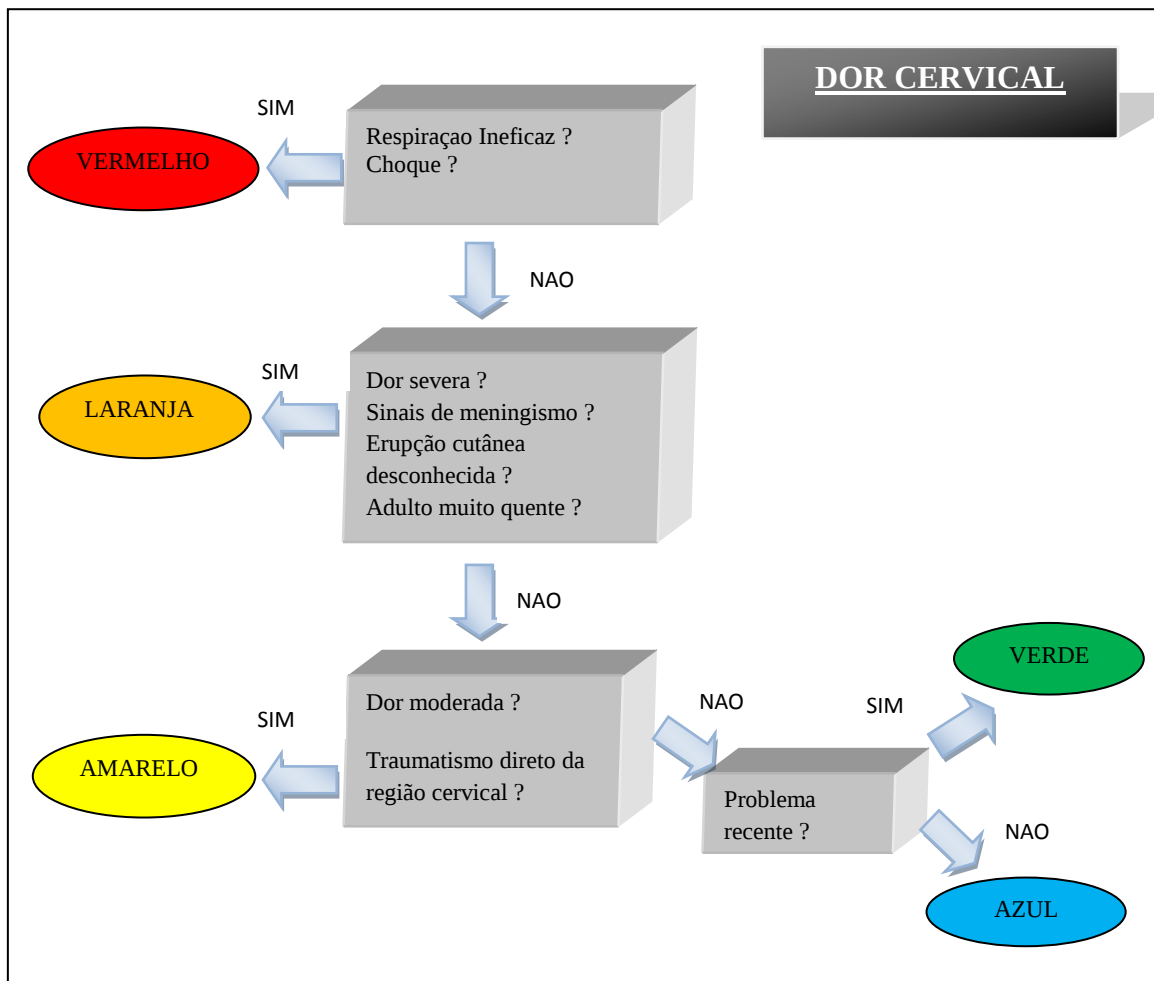


Figura 2.2 – Fluxograma orientador para a queixa principal tipo “dor cervical”, segundo o MTS

Fonte: Freitas (2002 apud Toledo, 2009)

2.2 Modelo de Priorização Australiano (Australian Triage Scale – ATS)

Na Austrália, os pacientes que procuravam assistência médica não eletiva no final dos anos 60 não eram sempre classificados. Aqueles que chegavam de ambulância eram priorizados, enquanto todos os demais eram avaliados por ordem de chegada (TOLEDO, 2009).

Segundo Esteves (2009), na década de 1970, nasceu o primeiro modelo de priorização no atendimento de pacientes. Foi no Box Hill Hospital, em Melbourne, e era conhecido como *Ipswich Triage Scale*. Esse modelo contava como uma escala de cinco prioridades, baseada em tempo e com identificação por adesivos na ficha de

atendimento médico. Porém, mais tarde, foi modificada com melhores descritores de urgência e necessidades de cuidados médicos. Somente nos anos de 1990, a *Ipswich Triage Scale* foi informatizada e testada em hospitais do continente australiano quanto à utilidade, aplicabilidade e validade.

Ainda de acordo com Esteves (2009), em 1993, *Australasian College of Emergency Medicine* adotou como modelo de priorização a *Ipswich Triage Scale*, porém, posteriormente, seu nome foi alterado para *National Triage Scale*, como parte das Políticas de triagem. Nos dias de hoje, esse modelo é mais conhecido como *Australasian Triage Scale (ATS)* e tem sido adotado pelas autoridades de saúde na Austrália e pelo *Australian Council on Health Care Standards* como base de avaliação do desempenho das unidades de urgência e emergência. Pioneiro e tem como característica principal a sistematização em tempos de espera, de acordo com a gravidade, o Modelo Australiano de Priorização tem cinco categorias de escala, listadas a seguir:

- Risco de vida imediato
- Risco de vida iminente
- Potencial Risco de vida
- Pacientes potencialmente graves
- Menos Urgentes

Este modelo pode ser utilizado para a classificação de risco de crianças, vítimas de trauma e distúrbios de comportamento. Além disso, sugere a reavaliação do paciente enquanto houver possibilidade de aparecimento de dados relevantes, enquanto o mesmo aguarda atendimento. A tabela 2.1 apresenta o modelo australiano, de acordo com as categorias, sua descrição e o tempo de avaliação.

Tabela 2.2 – Relação entre nível, descrição categoria e tempo de atendimento do ATS

Fonte: Australian College of Emergency Medicine (2001 apud Pires, 2003)

Categoria/Nível	Descrição da categoria	Avaliação
Categoria 1	Risco imediato à vida. Ex: parada cardio-respiratoria	Imediata
Categoria 2	Risco iminente à vida ou exigência de tratamento imediato. Ex : Insuficiência respiratória, dor torácica, perda de sangue importante	< 10 minutos
Categoria 3	Potencial ameaça à vida/urgência. Ex. Hipertensão severa, desidratação, convulsão.	30 minutos
Categoria 4	Situação de potencia urgência ou de complicações/gravidade importante. Ex. Aspiração de corpo estranho (sem obstrução de vias aéreas), dor abdominal não especifica.	60 minutos
Categoria 5	Menos urgente ou problemas clinico - administrativos. Ex: condições, necessidade de receita médica	120 minutos

2.3 Modelo de Priorização Canadense (Canadian Triage Acuity Scale - CTAS)

Proposto pela Associação Canadense de Medicina em Emergência, este modelo de priorização foi desenvolvido e validado em 1998 no Canadá. É composto por uma parte introdutória que aborda conceitos do tipo: triagem de pacientes que acessem as unidades de saúde destinadas ao atendimento de emergências em saúde; avaliação das condições do paciente e procedimentos a serem realizados por um profissional de saúde, geralmente um enfermeiro. Este modelo de priorização considera a classificação do pacientes numa escala ordinal que varia do número um ao número cinco (PIRES, 2003).

O nível um é destinado aos pacientes cuja intervenção deva ser imediata; já o nível dois refere-se a pacientes cujas condições de emergência requerem intervenções rápidas ou ações delegadas; o nível três reporta-se às condições de urgência que podem, potencialmente, progredir para um problema sério. Enquanto o nível quatro classifica os pacientes que se encontram em condições semi-urgentes relacionadas à idade ou à possibilidade de apresentarem deterioração ou complicação e, o último nível, o nível cinco, refere-se aos pacientes em condições não urgentes, agudas ou crônicas, cujas

intervenções podem sofrer atrasos ou cujos pacientes podem ser encaminhados a outras áreas (PIRES, 2003).

Segundo Beveridge (1998 apud Pires, 2003), o modelo preocupou-se com a implementação de três conceitos: utilidade, relevância e validade:

Utilidade: O modelo deve ser de fácil entendimento, melhorar a qualidade individual do atendimento e facilitar o funcionamento do serviço de emergência.

Relevância: A escala deve ter importância para os profissionais que estão utilizando-o, inclusive para os pacientes. Pode servir para diminuir a ansiedade do paciente e de sua família, assegurando que tempo de espera para o atendimento esteja adequado para sua condição clínica. O tempo de espera para que o paciente seja avaliado pelo médico, pode ser otimizado quando medidas de investigação ou de terapêutica baseadas em protocolos de evidências clínicas, forem iniciadas precocemente.

Validade: Deve não apenas ser consistente em sua aplicação, mas permitir visualização dos resultados como morbidade, mortalidade, hospitalização e utilização de recursos. A escala deve indicar resultados em grandes populações, pois muitas vezes, em casos isolados pode não apresentar bons resultados.

Tabela 2.3 – Relação entre nível, características e tempo de atendimento do CTAS

Fonte: Australian College of Emergency Medicine (2001 apud Pires, 2003)

Nível	Características	Tempo máximo previsto de espera do atendimento
1	Situações de risco à vida ou com sinais iminentes de risco de deterioração do quadro clínico. Exemplo: ataque cardíaco, insuficiência respiratória.	Imediata
2	Emergência. Condições que potencialmente ameaçam a vida ou requerem rápida intervenção. Exemplo: trauma craniano, alteração do estado mental.	≤15 minutos
3	Urgência. Condições que apresentam potencial para um problema sério. Exemplo: dispneia moderada, asma, convulsão.	≤30 minutos
4	Semi-urgência. Condições que apresentam potencial para complicações ou relacionadas à idade do paciente. Exemplo: dor abdominal, corpo estranho no olho.	≤60 minutos
5	Não urgência. Condições agudas não urgentes, ou problemas crônicos sem sinais de deterioração. Exemplo: dor de garganta, vômito e diarreia sem sinais de desidratação.	≤120 minutos

Para este modelo, recomenda-se aos profissionais da triagem que levem em consideração dados mais subjetivos: se o paciente parece doente, mas o profissional da triagem não tem tanta certeza do que realmente está acontecendo. Neste caso, recomenda-se que o paciente seja enquadrado no nível um ou dois. Este modelo, ainda conta com um manual em que estão definidas as condições clínicas presentes nos cinco níveis, e um resumo que permite rápido acesso a sinais e sintomas que devem ser observados. Faz considerações sobre o atendimento pediátrico, porém a associação desenvolveu uma nova escala exclusiva voltada para o atendimento de crianças nas emergências (PIRES, 2003).

A avaliação do atendimento realizada durante a triagem dos pacientes deve focalizar a queixa principal do paciente, baseando-se em dados subjetivos (início do problema, descrição da dor, fatores que agravam ou avaliam o problema e história prévia do mesmo problema) e dados objetivos (aparência física, resposta emocional, avaliação física e avaliação de sinais vitais, quando não houver premência de tempo

para o início do tratamento). Existe ainda a necessidade de dados complementares como o uso de medicações e alergias. Todo este processo deve durar de dois a cinco minutos (PIRES, 2003).

2.4 Modelo de Priorização Americano (Emergency Severity Index - ESI)

O Modelo Americano de priorização de atendimento de pacientes, usado em grande parte do território dos Estados Unidos, é baseado, assim como a maioria dos demais, em uma escala de cinco níveis. Os autores deste modelo, Richard Wuerz e David Eitel, acreditavam que o principal papel de um instrumento de triagem nos serviços de emergência na saúde é facilitar a priorização de pacientes, baseado na urgência da situação dos doentes. No momento da triagem, deve-se levantar uma questão do tipo: “Quem deve ser visto em primeiro lugar”? No entanto, para que o sistema não sobrecarregue, deve ser feita a seguinte questão: “quanto tempo, todos os pacientes podem esperar”? Neste sentido, o modelo foi desenvolvido em torno de um novo modelo conceitual de triagem. Atendendo o paciente certo para os recursos certos, no lugar certo e no momento certo (GYLBOY *et al.*, 2005).

Desenvolvido em 1998 com o objetivo de estratificar o risco dos pacientes, foi implantado em 1999 nos hospitais de ensino americanos e, após ser melhorado, foi implantado em mais cinco hospitais no ano de 2000. Este modelo estratifica o risco em cinco prioridades, estabelecendo tempo limite de espera para atendimento médico em cada nível de classificação. A escala trabalha com critérios clínicos guiados por um fluxograma (GYLBOY *et al.*, 2005).

Trata-se de um algoritmo de priorização baseado na gravidade e nos recursos necessários que oferece estratificação clínica dos pacientes em cinco grupos sendo o nível um o mais urgente e, o nível cinco, o menos urgente. Quando bem implementado, o ESI ajuda os serviços de emergência a rapidamente identificar os pacientes que necessitam de atenção imediata e melhor identifica os pacientes que podem ser encaminhados a cuidados em outros serviços (GYLBOY *et al.*, 2005).

A disposição do fluxograma do modelo americano de priorização de pacientes pode ser visualizada na figura:

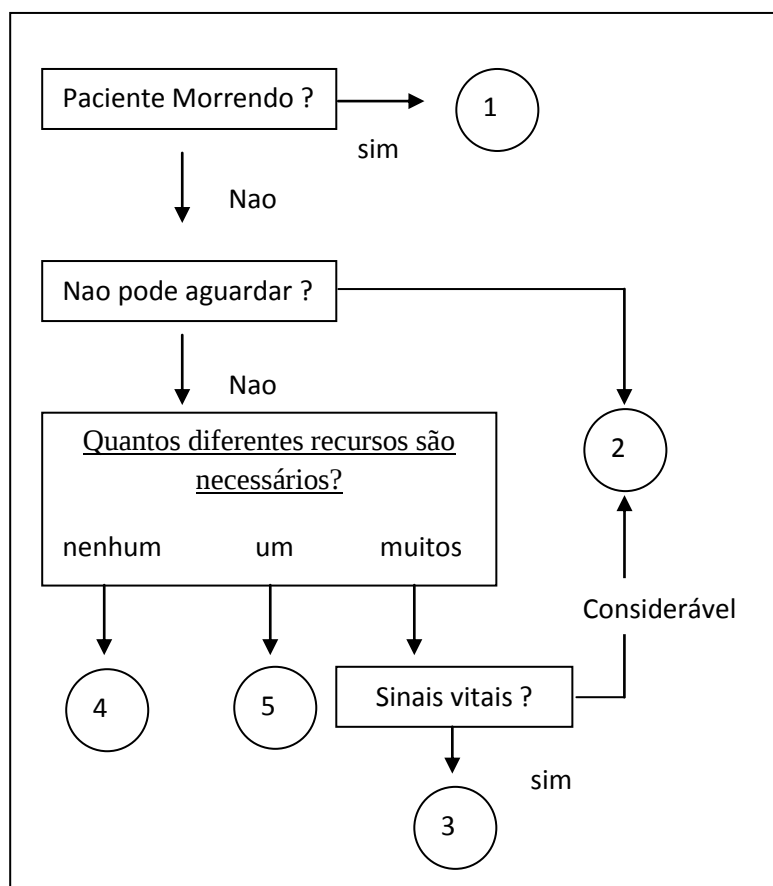


Figura 2.3 – Algoritmo disposição do fluxograma do ESI

Fonte: Gilboy (2005) (adaptado)

A Figura 2.3 ilustra o fluxograma a ser seguido pelo responsável da classificação dos pacientes. Através desse fluxograma, os pacientes são enquadrados em suas respectivas categorias. Nota-se que ESI engloba não somente o quadro clínico do paciente, mas também a quantidade de recursos necessários para o atendimento dos pacientes.

A ESI, assim como a CTAS, a ATS e o MTS também possui cinco níveis de classificação, que variam de um a cinco. Entretanto existem alguns pontos a serem destacados que a diferencia das demais. A principal diferença entre a ESI e as outras escalas é que estas consideram um tempo alvo máximo para o atendimento médico. Além disso, a ESI considera como imprescindíveis e determinantes os recursos disponíveis que a instituição dispõe de acordo com a potencial necessidade de utilização pelo paciente para a classificação do risco, o que remete à semelhança desse aspecto com a classificação de risco realizada em atendimento pré-hospitalar, em que, quando o número de vítimas excede a capacidade de recursos, deve-se priorizar o atendimento às

vítimas com melhor prognóstico, ou menos graves, ao passo que, quando o número de vítimas excede a capacidade de recursos, devem-se atender prioritariamente os clientes mais graves (TOLEDO, 2009; GILBOY ET *et al.*, 2005).

2.5 Comparações entre os modelos de priorização:

Os quatro modelos de priorização acima são os mais aceitos e mais avançados nos serviços de urgência e emergência dos principais hospitais do mundo. Apesar de serem muito parecidos e terem o mesmo propósito, possuem algumas particularidades que os diferenciam diante dos outros.

A tabela 2.3 ilustra as diferenças e características encontradas em cada um dos modelos apresentados.

Tabela 2.4 – Comparações entre os modelos ATS, CTAS, MTS e ESI

Fonte: GÓMEZ JIMÉNEZ (2003) (adaptado)

Características	Australiano	Canadense	Manchester	Americano
Escala de cinco Níveis	SIM	SIM	SIM	SIM
Utilização universal no país	SIM	SIM	SIM	NÃO
Baseado em categorias de sintomas	NÃO	NÃO	SIM	NÃO
Baseado em discriminante chave	SIM	NÃO	SIM	SIM
Baseado em algoritmos clínicos	NÃO	NÃO	SIM	SIM
Baseados em escalas de urgência pré-definidas	SIM	SIM	NÃO	NÃO
Formato eletrônico (Informatizado)	NÃO	NÃO	SIM	NÃO

Os quatro modelos possuem uma concepção sistêmica, ou seja, são utilizados por uma rede de serviços. De acordo com a tabela 2.4, nota-se em todos os modelos o uso de uma escala de cinco níveis. O único a apresentar um formato eletrônico é o modelo de Manchester, destacando-se entre os demais, já que o uso de sistemas de

informação traz diversos benefícios. Nota-se também que apenas o modelo de Manchester e o modelo Americano são baseados em algoritmos clínicos, possibilitando a padronização.

O MTS, além disso, é baseado em categorias de sintomas e não em escalas de urgência pré-definidas que podem induzir a diagnósticos, o que não é desejável em um modelo de priorização. A utilização universal no país corresponde ao uso do modelo nos principais hospitais onde o modelo foi desenvolvido (modelo Australiano, Austrália; modelo Canadense, Canadá; modelo de Manchester, Inglaterra e modelo Americano, Estados Unidos).

A tabela 2.4 apresenta uma comparação entre os tempos máximos previstos pelos modelos de priorização acima apresentados. Salientando que o modelo de priorização Americano não utiliza estes tempos.

Tabela 2.5 - Comparações entre os tempos máximos previstos em cada categoria do modelos MTS, CTAS e ATS

Fonte: O autor (2010)

Modelos	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5
MTS	Imediato	10 minutos	60 minutos	120 minutos	240 minutos
CTAS	Imediato	15 minutos	30 minutos	60 minutos	120 minutos
ATS	Imediato	10 minutos	30 minutos	60 minutos	120 minutos

De acordo com a tabela 2.5, para o nível 1, em todos os três modelos, o tempo previsto de atendimento é imediato. Ao nível 2, o MTS e o ATS convergem e preveem 10 minutos para o atendimento. Para os níveis 3, 4 e 5, os tempos dos CTAS e ATS são idênticos, 30, 60 e 120 minutos, respectivamente. No MTS, estes tempos são de 60, 120 e 240 minutos, respectivamente.

3. TEORIA DAS FILAS

A Teoria das Filas é uma ferramenta matemática que através de análises quantitativas visa tratar de processos aleatórios, como atrasos e congestionamentos, e tem como objetivo otimizar o desempenho do sistema e reduzir seus custos operacionais.

Segundo Hillier & Lieberman (2006), a teoria das filas é o estudo da espera em diversas formas, como comprar ingresso numa sessão de cinema, fazer um depósito bancário, comprar um sanduíche em uma lanchonete ou aguardar para ser atendido num hospital. Essa teoria usa modelos de filas para representar os diversos tipos de sistemas de filas (sistemas que envolvem filas do mesmo tipo) que surgem na prática. As fórmulas para cada modelo indicam como o sistema de filas correspondente deve funcionar, inclusive o tempo de espera médio que ocorrerá.

Ainda de acordo com Hillier & Lieberman (2006), os modelos de filas são extremamente úteis para determinar como operar um sistema de filas da forma mais eficiente possível. A capacidade de atendimento em excesso resulta em custos demasiados, enquanto que o atendimento insuficiente gera uma espera excessiva do cliente resultando em diversas consequências indesejáveis para os clientes e para a organização como um todo. Estes modelos permitem encontrar um equilíbrio apropriado entre custo de serviço e o tempo de espera.

Fogliatti & Mattos (2007, p.1), escrevem o seguinte:

A teoria das filas consiste na modelagem analítica de processos ou sistemas que resultam em espera e tem como objetivo determinar e avaliar quantidades, denominadas medidas de desempenho, que expressam a produtividade/operacionalidade desses processos. Entre essas medidas, podem-se citar: número de elementos na fila e tempo de espera pelo atendimento tempo ocioso dos prestadores de serviço.

3.1. Estrutura básica dos modelos de filas

3.1.1 Processos de filas básico

A maioria dos modelos de filas utiliza um processo básico da seguinte maneira: Através de uma fonte de entrada, clientes que desejam atendimento chegam ao longo do tempo, entrando no sistema de filas e pegando uma senha. De acordo com alguma regra conhecida como disciplina da fila, um determinado cliente em algum momento é atendido. Este atendimento é realizado para o cliente pelo mecanismo de atendimento. Após esse processo, o cliente atendido deixa o sistema de filas (HILLIER & LIEBERMAN, 2006).

A figura 3.1 representa o processo acima descrito:

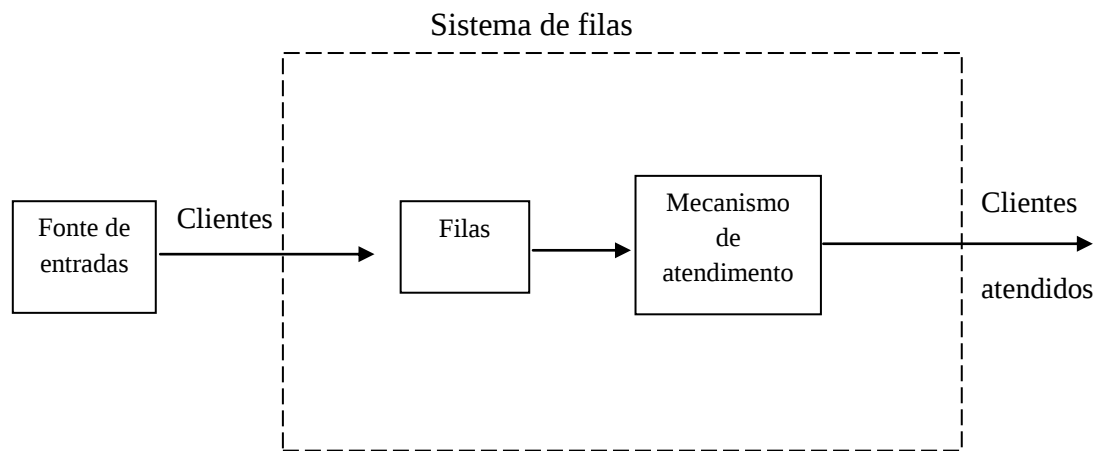


Figura 3.1 – Representação básica de um sistema de filas

Fonte: Hillier & Lieberman (2006)

3.1.2 Fonte de entradas (População Solicitante)

Segundo Hillier & Lieberman (2006), uma das características da fonte de entradas é o seu tamanho, que diz respeito ao número total de clientes que possam precisar de um atendimento, ou seja, corresponde ao número total de possíveis clientes diferentes e pode ser de tamanho finito ou infinito. População Solicitante representa a população de onde provêm as chegadas dos clientes. Este processo de chegada estabelece um padrão estatístico de como os clientes chegam ao sistema.

De acordo com Costa (2005), a distribuição mais comum para o processo de chegada de clientes é a distribuição de Poisson, onde os tempos entre chegadas são exponencialmente distribuídos. Comportando-se como em um comportamento estocástico, pois o tempo entre as chegadas dos clientes é uma variável aleatória.

Outras hipóteses de comportamento entre as chegadas de clientes podem ser assumidas, como Erlang, hiperexpoencial e arbitrária.

Comportamentos incomuns por parte dos clientes que chegam ao sistema devem ser mencionados, como é o caso do cliente decepcionado, ou seja, é o cliente que se recusa a entrar no sistema de filas devido ao tamanho da fila ou ainda, depois de entrar no sistema de filas pode desistir e sair definitivamente devido a uma grande espera.

3.1.3 Tempo de serviço

Tempo de serviço é o intervalo de tempo decorrido entre o início até o término do atendimento do cliente. Costuma também ser chamado de tempo de permanência. O tempo de serviço pode ser dependente ou independente do estado. Dizemos que é dependente do estado quando, por exemplo, um atendente compromete o tempo de serviço por algum motivo, como o número de clientes na fila. Dizemos que é independente do estado quando o tempo de serviço em nada varia, ou seja, o tempo de atendimento é distribuído de forma idêntica (COSTA, 2005).

3.1.4 Fila

A fila corresponde ao local onde os clientes aguardam antes de ser atendidos. De acordo com o número máximo de clientes que a fila pode suportar, podem ser finitas ou infinitas. Para fins de facilidade de cálculos, a maioria dos modelos de filas considera como padrão a fila sendo infinita, a hipótese de uma fila finita seria um fator complicador na análise (HILLIER & LIEBERMAN, 2006).

3.1.5 Disciplina das filas

A disciplina das filas é a ordem na qual os clientes serão selecionados ou priorizados para o atendimento. Existem diversos casos de disciplinas das filas, como exemplo pode-se citar a FIFO (*First in, First out*) ou primeiro a entrar, primeiro a sair. O primeiro cliente na fila deverá ser atendido em primeiro lugar, na frente dos demais. É respeitada uma ordem de chegada. Em geral, num sistema de filas, adota-se esse critério, ao menos que seja combinada de alguma outra forma. Outro exemplo é a LIFO (*Last in, First out*) ou último a chegar, primeiro a sair.

Podem ocorrer também filas com prioridades, onde alguns clientes têm prioridades de atendimento em relação a outros clientes. Segundo Hillier & Lieberman

(2006), existem dois casos desse tipo: preemptivo e não-preemptivo. O caso preemptivo é aquele em que o cliente com prioridade será atendido no momento da sua chegada, independentemente de outro cliente estar sendo atendido na hora. Neste caso, ocorre que o atendimento atual será interrompido para o atendimento do cliente com prioridade, sendo retomado ao final do mesmo. Como ocorre nos serviços de emergências da maioria dos hospitais. Clientes que chegam num estado de saúde mais grave devem ser atendidos prioritariamente em relação aos outros pacientes que já se encontram na fila. Sistemas preemptivos-retomados se caracterizam quando o atendimento para um cliente preterido é retomado no ponto onde foi interrompido, e sistemas preemptivos-repetitivos são aqueles onde o atendimento tem de começar do início novamente. O caso não-preemptivo é aquele que ocorre quando o atendimento que está sendo realizado naquele momento não poderá ser interrompido, o cliente que chega com maior prioridade deverá ir para o início da fila e esperar até que o atendimento atual de menor prioridade seja finalizado.

3.1.6 Mecanismo de atendimento ou canais de serviço

O mecanismo de atendimento é formado por uma ou mais instalações de atendimento. Cada instalação de atendimento pode conter um ou mais servidores ou canais de atendimento em paralelo. Um servidor é a unidade de produção responsável pela realização do atendimento. Em um modelo de filas, deve-se especificar a disposição das instalações e o número de atendentes em cada uma delas, e também a distribuição probabilística de tempos de atendimento para cada atendente e, se possível, para tipos diferentes de clientes. Entretanto, na maioria dos modelos de filas se parte do pressuposto que os tempos de atendimentos para diferentes clientes são iguais. A distribuição de tempo de atendimento, na maioria dos modelos, segue uma distribuição exponencial, porém em alguns outros modelos, pode-se usar a distribuição degenerada (tempo de atendimento constante) e a distribuição de Erlang (HILLIER & LIEBERMAN, 2006).

3.1.7 Processo de filas elementar

O processo de filas elementar, mais frequentemente, parte da seguinte situação: existe uma fila de espera única que pode estar vazia ou não. Essa fila se encontra em frente de uma única instalação de atendimento, que pode conter um ou mais atendentes.

Cada cliente que chega através da fonte de entradas é atendido por um atendente ou servidor.

A figura 3.2 ilustra essa situação:

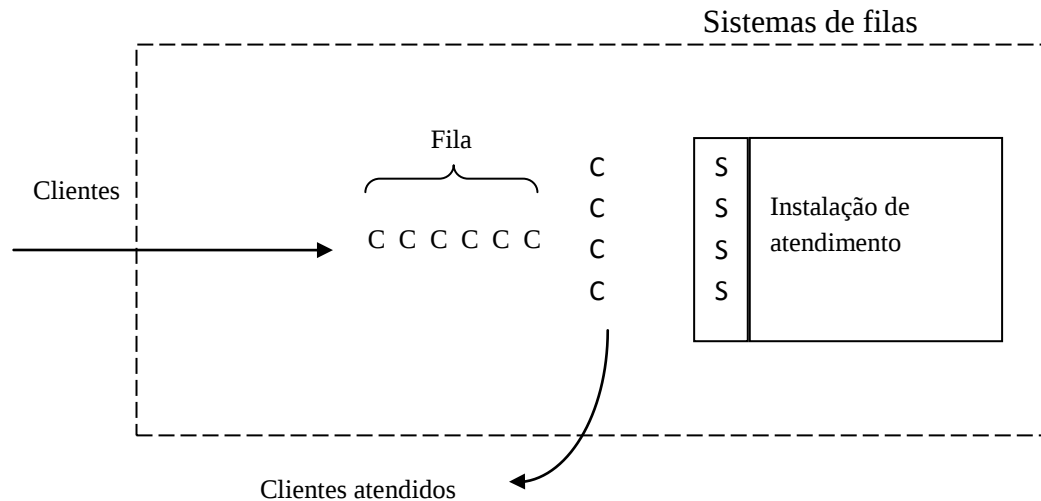


Figura 3.2 – Representação do processo de filas elementar

Fonte: Hillier & Lieberman (2006)

A representação da figura 3.2 corresponde à maioria das situações onde haja um sistema de filas, com uma fonte de entrada, um fila e um servidor. Porém, existem ainda outras representações que possa melhor se adaptar a uma determinada situação. Neste trabalho, foi focado nesta representação, já que melhor representa o objeto de estudo. A fonte de entrada gera clientes na forma de casos de emergências. Uma sala de emergência representa uma instalação de atendimento e os médicos são representados pelo atendente.

3.1.8 Notação

Encontram-se na literatura algumas diferenças entre as notações usadas de autor para autor. Foram usadas, neste trabalho, as notações utilizadas por Hillier & Lieberman (2006), da seguinte maneira:

- Estado do sistema = número de clientes no sistema de filas.
- Comprimento da fila = Número de clientes aguardando que o atendimento se inicie ou ainda, estado do sistema *menos* número de clientes que estão sendo atendidos.

- $N(t)$ = número de clientes no sistema de filas no instante t ($t \geq 0$).
- $P_n(t)$ = probabilidade de exatamente n clientes se encontrarem no sistema de filas no instante t , dado o número no instante 0.
- s = número de atendentes (canais de atendimento paralelos) no sistema de filas.
- λ_n = taxa média de chegada (número de chegadas esperado por unidade de tempo) de novos clientes quando n clientes se encontram no sistema.
- μ_n = taxa média de atendimento para o sistema global (número de clientes esperado completando o atendimento por unidade de tempo) quando n clientes se encontram no sistema. μ_n representa a taxa *combinada* na qual todos os atendentes *ocupados* (aqueles que se encontram atendendo clientes) completam o atendimento.

Quando λ_n for uma constante para todo n , essa constante é representada por λ . Quando a taxa média de atendimento por atendente ocupado for uma constante para todo $n \geq 1$, essa constante é representada por μ . Nesse caso, $\mu_n = s\mu$, quando $n \geq s$, isto é, quando todos os s atendentes estiverem ocupados. Sob essas condições, $1/\lambda$ e $1/\mu$ são, respectivamente, o tempo esperado entre atendimentos e o tempo de atendimento esperado. Da mesma forma, $\rho = \lambda/(s\mu)$ é o fator de utilização para a instalação de atendimento, isto é, a fração de tempo esperada em que atendentes individuais se encontram ocupados, pois $\lambda/(s\mu)$ representa a fração da capacidade de atendimento ($s\mu$) do sistema que está sendo utilizada em média pelos clientes que chegam (λ).

De acordo do Hillier & Lieberman (2006), quando um sistema de filas começa a operar recentemente, este passa por um processo de transitoriedade, denominado de condição transitória. É quando o estado do sistema (números de clientes no sistema) é afetado pelo tempo que passou e pelo estado inicial. Após decorrido algum tempo, um condição de estabilidade se concretiza. Denomina-se essa situação, quando não há mais dependência com o estado inicial nem com o tempo decorrido, de condição de estado favorável, na qual a distribuição probabilística do estado do sistema permanece a mesma (a distribuição de estado estável ou estacionário) ao longo do tempo. As

notações a seguir partem do pressuposto que a condição em que o sistema se encontra seja uma condição de estado favorável.

Uma condição transitória exigiria um conhecimento analítico além das fronteiras desse trabalho e além do que, o estudo de caso na qual este trabalho teve a intenção de mostrar, remonta a um a estado de condição favorável.

- P_n = Probabilidade de exatamente n clientes se encontrarem no sistema de filas
- L = Número de clientes esperado no sistema de filas = $\sum_{n=0}^{\infty} nP_n$
- L_q = Comprimento esperado da fila (exclui clientes que estão sendo atendidos) = $\sum_{n=s}^{\infty} (n - s)P_n$
- W = Tempo de espera no sistema (inclui o tempo de atendimento) para cada cliente individual, $W = E(W)$.
- W_q = Tempo de espera na fila (exclui o tempo de atendimento) para cada cliente individual, $W_q = E(W_q)$.

3.1.8.1 Relações entre L , W , L_q e W_q e fórmula de Little

Supondo que λ_n seja uma constante λ para todo n , num processo de filas em estado favorável, chega-se a seguinte conclusão:

$$L = \lambda W, \quad (1)$$

Esta fórmula, mais conhecida como fórmula de Little, não leva em consideração as distribuições de tempos entre chegadas nem tempos de atendimento.

Além disso, matematicamente, pode-se dizer também que:

$$L_q = \lambda W_q \quad (2)$$

Supondo agora que λ_n não forem iguais, então substitui-se λ pela taxa média de chegada a longo prazo, representado por $\bar{\lambda}$.

Supondo que o tempo médio de atendimento seja constante, $1/\mu$ para todo $n \geq 1$, conclui-se que:

$$W = W_q + \frac{1}{\mu} \quad (3)$$

3.1.9 Medidas de desempenho

Fogliatti & Matos (2007) destacam que quando se utiliza da teoria das filas para análise e otimização de um processo, procura-se obter ganhos nas performances de determinados parâmetros de desempenho ou medidas de desempenho, como: número de clientes esperado no sistema de filas (L); comprimento esperado da fila (L_q); tempo de espera no sistema para cada cliente individual (W) e/ou tempo de espera na fila (exclui o tempo de atendimento) para cada cliente individual (W_q).

3.2. Modelos de filas baseados no processo de nascimento-e-morte

Segundo Hillier & Lieberman (2006), no contexto da teoria das filas, entende-se que nascimento é a chegada de um novo cliente na fila, enquanto que morte representa a partida de um cliente atendido. Desta forma, surge o conceito do processo de nascimento-e-morte, representado pela entrada, clientes que chegam, e pela saída, cliente atendidos que saem do sistema de filas. O processo de nascimento-e-morte descreve probabilisticamente como $N(t)$, o número de clientes no sistema de filas, varia de acordo com t , quando t aumenta, para $t \geq 0$. Nascimentos e mortes individuais ocorrem aleatoriamente e suas taxas médias de ocorrência dependem apenas do estado momentâneo do sistema. Três hipóteses são consideradas no processo de nascimento-e-morte:

- Hipótese 1: Dado $N(t) = n$, a distribuição probabilística atual do tempo remanescente até o próximo nascimento (chegada) é exponencial com parâmetro $\lambda_n = (n = 0, 1, 2, \dots)$.
- Hipótese 2: Dado $N(t) = n$, a distribuição probabilística atual do tempo remanescente até a próximo morte (término do atendimento) é exponencial com parâmetro $\mu_n (n = 1, 2, \dots)$.
- Hipótese 3: A variável aleatória da hipótese 1 (o tempo remanescente até o próximo nascimento) e a variável aleatória da hipótese 2 (o tempo remanescente até a próxima morte) são mutuamente independentes.

Diversos modelos são propostos pela teoria das filas, de acordo com as particularidades presentes em cada tipo de situação. Neste trabalho, será apresentado o modelo M/M/s que servirá de base, assim também como os modelos de filas de disciplina de prioridades, pois, segundo Hillier & Lieberman (2006), são os que mais se aproximam das situações que envolvem um serviço hospitalar de emergência, objeto de estudo deste trabalho.

3.2.1 Modelo M/M/s

Hillier & Lieberman (2006, p. 21) exprimem a ideia do modelo M/M/s da seguinte forma:

“O modelo M/M/s parte do pressuposto de que todos os tempos entre atendimentos sejam distribuídos de forma independente e idêntica de acordo com uma distribuição exponencial (isto é, o processo de entrada é de Poisson), que todos os tempos de atendimento sejam distribuídos de forma independente e idêntica de acordo com outra distribuição exponencial e que o número de atendentes seja s (qualquer inteiro positivo). Consequentemente, esse modelo é simplesmente o caso especial do processo de nascimento-e-morte em que a taxa média de chegada e a taxa média de atendimento por atendente ocupado do sistema de filas são constantes (λ e μ , respectivamente) independente do estado do sistema.

Os diagramas de taxas para ($s = 1$) e ($s > 1$) são representados de acordo com as figuras 3.3 e 3.4, respectivamente.

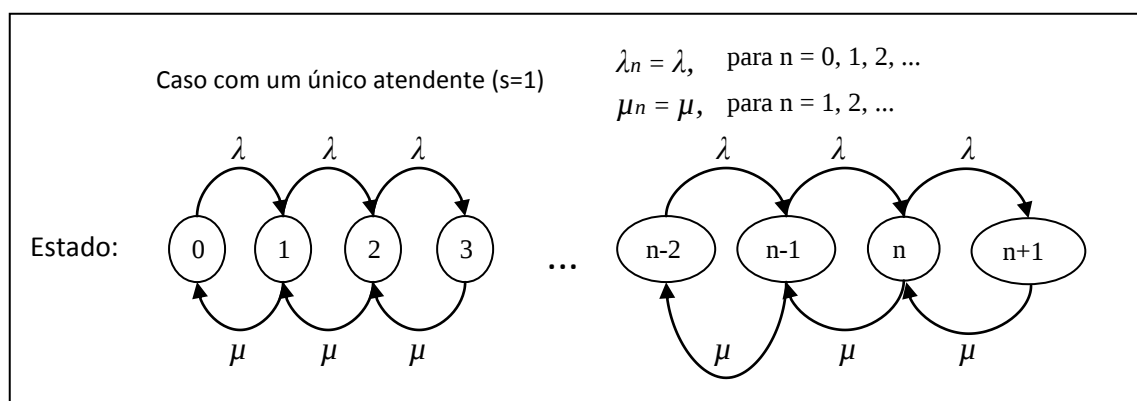
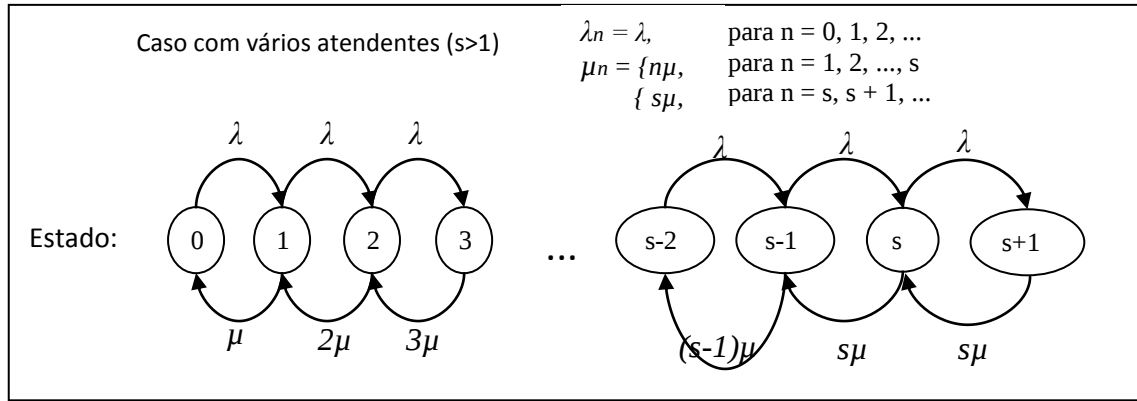


Figura 3.3 - Diagrama de taxas para o modelo M/M/s ($s = 1$)

Fonte: Hillier & Lieberman (2006)

Figura 3.4 - Diagrama de taxas para o modelo M/M/s ($s > 1$)

Fonte: Hillier & Lieberman (2006)

3.2.1.1 Caso com um único atendente ($s = 1$) (M/M/1)

Segundo Fogliatti & Mattos (2007), neste modelo, as chegadas e o atendimento se comportam como em um processo de nascimento e morte, a distribuição dos tempos entre as chegadas e a distribuição de probabilidade do tempo de serviço são exponenciais. Tanto a taxa de chegadas ao sistema quanto a taxa de atendimento são constantes e há apenas um atendente. A regra de disciplina neste modelo é a FIFO.

Em regime estacionário, quando há apenas um atendente ($s = 1$), chega-se às seguintes conclusões:

- Fator de utilização do sistema:

$$\rho = \frac{\lambda}{\mu} \quad (4)$$

- Os fatores C_n para o processo de nascimento-e-morte ou a probabilidade de ser mais do que n elementos do sistema:

$$C_n = \left(\frac{\lambda}{\mu}\right)^n = \rho^n, \quad \text{para } n = 0, 1, 2, \dots \quad (5)$$

- Probabilidade de se ter zero clientes no sistema:

$$P_0 = 1 - \rho \quad (6)$$

- Probabilidade de exatamente n clientes se encontrarem no sistema de filas:

$$P_n = (1 - \rho)\rho^n, \quad \text{para } n = 0, 1, 2, \dots \quad (7)$$

- Probabilidade do tempo de espera ser maior do que um tempo $t > 0$:

$$P\{W_q > t\} = \rho e^{-\mu(1-\rho)t} \quad (8)$$

- Número de clientes esperado no sistema:

$$L = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda} \quad (9)$$

- Comprimento esperado da fila:

$$L_q = \frac{(\lambda)^2}{\mu(\mu - \lambda)} \quad (10)$$

- Tempo de espera no sistema:

$$W = \frac{1}{\mu - \lambda} \quad (11)$$

- Tempo de espera na fila:

$$W_q = E(W_q) = \frac{\lambda}{\mu(\mu - \lambda)} \quad (12)$$

3.2.1.2 Caso com vários atendentes ($s > 1$) (M/M/s)

Quando há vários atendentes ($s > 1$), chega-se às seguintes conclusões:

- Fator de utilização do sistema:

$$\rho = \frac{\lambda}{s\mu} \quad (13)$$

- Os fatores C_n para o processo de nascimento-e-morte ou a probabilidade de ser mais do que n elementos do sistema:

$$C_n = \begin{cases} \frac{(\frac{\lambda}{\mu})^n}{n!}, & \text{para } n = 1, 2, \dots, s \\ \frac{(\frac{\lambda}{\mu})^n}{s!s^{n-s}}, & \text{para } n = s, s+1, \dots \end{cases} \quad (14)$$

- Se $\lambda < s\mu$, a probabilidade de se ter zero clientes no sistema é dada por:

$$P_o = \frac{1}{\left[\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{1 - \{\lambda/(s\mu)\}} \right]} \quad (15)$$

- Comprimento esperado da fila:

$$L_q = \frac{P_o(\lambda/\mu)^s \rho}{s!(1-\rho)^2} \quad (16)$$

- Tempo de espera na fila:

$$W_q = \frac{L_q}{\lambda} \quad (17)$$

- Tempo de espera no sistema:

$$W = W_q + \frac{1}{\mu} \quad (18)$$

- Número de clientes esperado no sistema:

$$L = L_q + \frac{\lambda}{\mu} \quad (19)$$

- Distribuição probabilística dos tempos de espera:

$$P\{W > t\} = e^{-\mu t \left[\frac{1+P_o(\lambda/\mu)^s}{s!(1-\rho)} \left(\frac{1 - e^{-\mu t(s-1-(\lambda/\mu))}}{s-1-(\lambda/\mu)} \right) \right]} \quad (20)$$

e,

$$P\{W_q > t\} = (1 - P\{W_q = 0\}) e^{-s\mu(1-\rho)t} \quad (21)$$

sendo que,

$$P \{W_q = 0\} = \sum_{n=0}^{s-1} P_n \quad (22)$$

Caso $\lambda \geq s\mu$, isto quer dizer que a taxa média de chegada excede a taxa média de atendimento máximo, fazendo com que a fila cresça sem limites, e as fórmulas acima descritas que servem para um estado estacionário não mais se aplicam.

3.2.2 Modelos de filas de disciplinas de prioridades

Nos modelos de filas de disciplinas de prioridades, a disciplina de filas é baseada em um sistema de prioridades. Os clientes que aguardam nas filas para serem atendidos são selecionados de acordo com uma ordem baseada nas prioridades que lhes foram atribuídas. Portanto, certos clientes têm prioridades em relação a outros. A utilização de modelos de prioridades geralmente fornece um refinamento bem adequado quando comparados com outros modelos de filas mais usuais (HILLIER & LIEBERMAN, 2006).

De acordo com Hillier & Lieberman (2006), existem dois modelos de filas de disciplina de prioridades e ambos são muito parecidos; o que os diferencia é se as prioridades são não-preemptivas ou preemptivas. Em ambos os modelos pressupõem-se que existam N classes de prioridades, sendo a classe 1 com prioridade mais alta, enquanto que a classe N tem a prioridade mais baixa. Toda vez que um atendente ficar livre para atender um novo cliente, deverá atender o primeiro da classe de prioridade mais alta. E dentro de uma mesma classe, os clientes são selecionados de acordo por ordem de chegada. É considerado que o processo de chegada segue uma entrada de Poisson e tempos de atendimentos são exponenciais para classe de prioridades. Considera-se também uma hipótese bastante restritiva em que o tempo de atendimento esperado seja o mesmo para todas as classes de prioridades. Na prática, é possível que essa hipótese seja violada. No caso único de apenas um atendente ($s = 1$) e esta hipótese seja violada, é possível permitir tempos de atendimento esperado diferentes obtendo resultados úteis. Entretanto, as taxa média de chegada são diferentes entre as diversas classes de prioridades.

Ainda de acordo com Hillier & Lieberman (2006), ambos os modelos são idênticos ao modelo M/M/s, com a exceção da ordem pela qual os clientes são atendidos.

Consequentemente, a distribuição de estado estável do modelo M/M/s também se aplica aos modelos de filas de disciplina de prioridades. As fórmulas de L e L_q são as mesmas, assim como os resultados esperados de tempo de espera (W e W_q) para qualquer cliente que seja escolhido aleatoriamente. O objetivo é melhorar as medidas de desempenho para cada uma das classes de prioridades mais altas à custa de desempenho para as classes de prioridades mais baixas.

3.2.2.1 Com prioridades não-preemptivas

Seja W_k o tempo de espera previsto no estado estacionário do sistema para um cliente da classe de prioridades k ; s igual ao número de atendentes; μ igual à taxa média de atendimento por atendente ocupado e λ_i igual à taxa média de chegada por classe de prioridade i . Desta forma, chega-se às seguintes conclusões:

$$W_k = \frac{1}{AB_{k-1}B_k} + \frac{1}{\mu}, \text{ para } k = 1, 2, \dots, N, \quad (23)$$

em que:

$$A = s! \frac{s\mu - \lambda}{r^s} \sum_{j=0}^{s-1} \frac{r^j}{j!} + s\mu, \quad (24)$$

$$B_0 = 1, \quad (25)$$

$$B_k = 1 - \frac{\sum_{i=1}^k \lambda_i}{s\mu} \quad (26)$$

$$\lambda = \sum_{i=1}^N \lambda_i \quad (27)$$

$$r = \frac{\lambda}{\mu} \quad (28)$$

Através da fórmula de Little, o número esperado no estado estável de clientes da classe de prioridades k (L_k) o sistema de filas é representado por:

$$L_k = \lambda_k W_k, \text{ para } k = 1, 2, \dots, N. \quad (29)$$

O tempo de espera previsto na fila (excluindo o tempo de atendimento) para a classe de prioridades k é determinado subtraindo-se $1/\mu$ de W_k ; o comprimento esperado da fila correspondente é novamente obtido multiplicando por λ_k .

3.2.2.2 Com prioridades preemptivas

- Tempo de espera total previsto no sistema (incluindo o tempo total de atendimento), quando s é igual a um ($s = 1$).

$$W_k = \frac{1/\mu}{B_{k-1}B_k}, \text{ para } k = 1, 2, \dots, N. \quad (30)$$

Para ($s > 1$), ainda é válida a relação $L_k = \lambda_k W_k$, para $k = 1, 2, \dots, N$, assim como o tempo de atendimento total previsto para um cliente selecionado aleatoriamente continua sendo $1/\mu$. E os resultados correspondentes para a fila também podem ser obtidos de W_k e L_k como no caso com prioridades não-preemptivas.

4. Estudo de Caso

4.1. O Hospital da Restauração de Pernambuco (HR)

O serviço de urgência e emergência em estudo é o do Hospital da Restauração de Pernambuco, situado na região metropolitana da cidade do Recife. Sua localização é central e bastante estratégica, sendo o hospital de referência para a própria cidade e todos os municípios vizinhos. O serviço de urgência e emergência do HR funciona vinte e quatro horas, sete dias por semana e faz parte da rede do Sistema Único de Saúde.

Com mais de 530 leitos, o HR se caracteriza como um hospital de grande porte e pertence à Secretaria de Saúde do Estado de Pernambuco. Ao longo dos anos, o HR vem demonstrando como uma marcante vocação para casos de alta complexidade, emergência clínica, politraumatismos e neurocirurgia.

O serviço de urgência e emergência do HR é baseado no Sistema de Triagem de Manchester, classificando os pacientes que procuram este tipo de atendimento em cinco categorias, de acordo com o tempo de atendimento e cor específica. Foi verificado que sempre que um paciente classificado na categoria *Emergência* chegasse no sistema de filas da emergência do HR, este seria atendido rapidamente, mesmo se um médico estivesse atendendo outro paciente de prioridade menor, comportando-se como um modelo de prioridades preemptivas. Entretanto, não foi verificado este comportamento nas outras categorias, comportando-se como um modelo de prioridades não-preemptivas. Desta forma, considerou-se o modelo de filas do HR como sendo o de prioridades não-preemptivas, já que apenas a categoria *Emergência* segue prioridades preemptivas.

4.2. As Unidades de Pronto-Atendimento (UPAs)

As Unidades de Pronto-atendimento auxiliam os principais hospitais tratando de casos de complexidade intermediária, desafogando os hospitais. Geralmente são espalhados em lugares estratégicos, longe dos hospitais, no intuito de descentralizar os serviços de saúde da rede do SUS. Quando bem espalhados, são de significativa importância para diminuir a demanda dos serviços de urgência e emergência dos hospitais, contribuindo para um serviço prestado de maior qualidade. Desde que foram implantadas na cidade do Recife e nos diversos municípios vizinhos, as UPAs vêm

desafogando o serviço de emergência do HR e melhorando as condições para os pacientes.

As UPAs podem ser classificadas em três diferentes portes, dependendo do tamanho da população a ser coberta, a capacidade instalada (área física, número de leitos disponíveis, recursos humanos e capacidade diária de atendimento médico. Para cada porte há um incentivo financeiro diferente de investimento para a implementação das mesmas além do custeio mensal (MINISTÉRIO DA SAÚDE, 2010)

4.3. Coleta e análise dos dados

O trabalho trata especialidade de clínico geral, devido ao fato de que todos os pacientes que buscam o serviço de emergência do HR passam por este profissional no início do atendimento, para, então, serem encaminhados à realização de exames, marcação de consultas, atendimentos a outros médicos mais especializados, cirurgias, ou mesmo liberados com alta médica, além da ocorrência de reclamações por parte dos pacientes relacionados à demora e à qualidade do atendimento nesta parte do serviço prestado.

A recepção dos pacientes funciona da seguinte maneira: o paciente dá entrada no serviço de urgência e emergência do HR, recebendo uma senha de atendimento. Este é classificado por uma enfermeira de acordo com a classificação de Triagem de Manchester. Em seguida, o paciente é classificado numa das cinco categorias com a pulseira de uma determinada cor e aguarda na fila para ser atendido pelo clínico geral.

Os dados foram referentes ao período de 12 a 18 de novembro do ano de 2010, coletados pelo Sistema de Informação do HR, a partir do registro de entrada dos pacientes no HR e entregues em forma de relatórios impressos ao autor. Foram coletados dados referentes a 381 pacientes, sem distinções de classe, idade, município de origem ou outras formas do gênero.

Tabela 4.1 – Quantidade de pacientes classificados e atendidos por dia

Fonte: Sistema de Informação do HR. (2010)

Data	Dia	Pacientes classificados	Pacientes atendidos
12/11/2010	Sexta-feira	57	54
13/11/2010	Sábado	57	56
14/11/2010	Domingo	52	52
15/11/2010	Segunda-feira	59	59
16/11/2010	Terça-feira	57	57
17/11/2010	Quarta-feira	55	53
18/11/2010	Quinta-feira	51	49
Total		388	380
Média de pacientes por dia da semana		55,43	54,29

A tabela 4.1 mostra a quantidade diária de pacientes atendidos entre os dias 12 à 18 de novembro, assim também como o total desses pacientes na semana considerada e uma média dos atendimentos por dia. Verifica-se que ao total foram atendidos 380 pacientes, enquadrados em uma das cinco categorias do modelo de Manchester. A média foi de 54,29 pacientes atendidos por dia.

Em alguns dias na coleta percebe-se que o número de pacientes classificados é ligeiramente maior que o número de pacientes atendidos. Este fato deve-se a desistência de alguns pacientes depois que foram classificados na Triagem de Manchester ou até mesmo, após serem classificados, foram encaminhados às UPAs para serem atendidos.

A figura 4.1 ilustra a quantidade de pacientes atendidos no período de uma semana em cada uma das cinco categorias (emergência, muito urgente, urgente, pouco urgente e não urgente), assim como a porcentagem relativa sobre o total no mesmo dia.

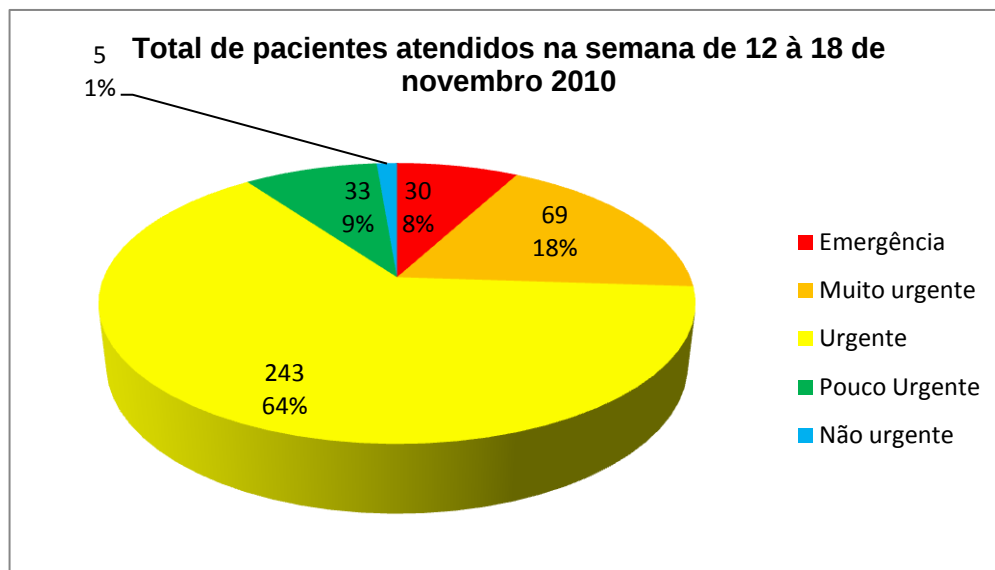


Figura 4.1 – Total de pacientes atendidos durante o período de 12/11/2010 à 18/11/2010 classificados de acordo com sua gravidade e cor do Sistema de Triagem de Manchester

Fonte: Sistema de Informação do HR (2010)

No gráfico da figura 4.1 percebe-se que a grande maioria, 243 pacientes atendidos, representando 64% do total (380) de pacientes, é classificada na categoria *Urgente* (indicado pela cor amarela). A categoria *Muito urgente* (em laranja) conta com 69 pacientes atendidos, representando 18% do total. Em seguida a categoria *Pouco urgente* (em verde) representa 9% do total com 33 pacientes, enquanto que na categoria *Emergência* (em vermelho) foram atendidos 30 pacientes, ou seja, 8% do total. Por último, a categoria *Não urgente* (em azul) foi a menos representativa, com apenas cinco pacientes atendidos nesse período de tempo, apenas 1% do total.

O gráfico da figura 4.2 ilustra os horários de chegada e a quantidade de pacientes atendidos durante o decorrer das horas na semana estudada.

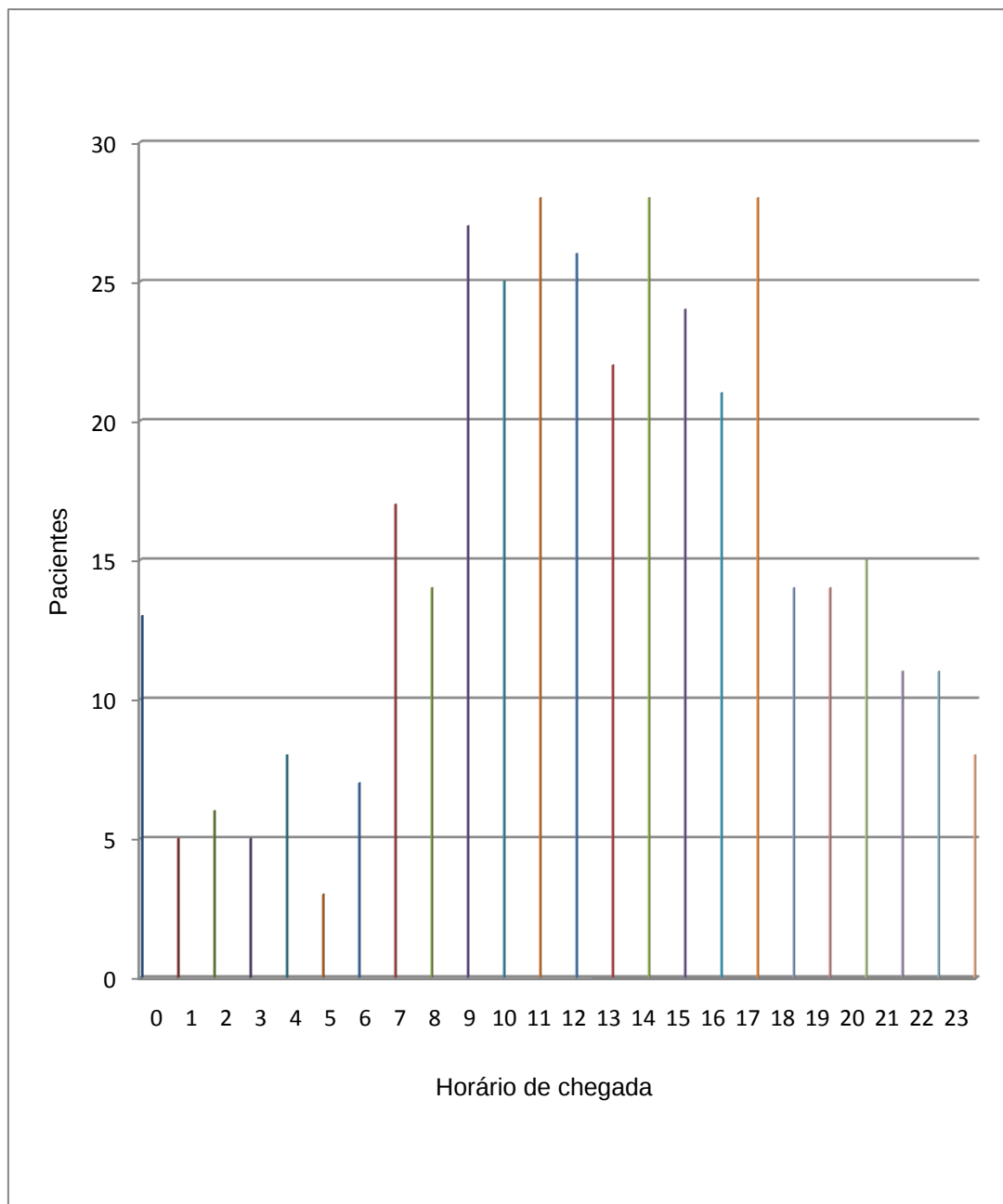


Figura 4.2 – Total de pacientes na semana x hora do dia

Fonte: Sistema de Informação do HR (2010)

De acordo com a Figura 4.2 nota-se que no período entre 9 horas até às 18 horas ocorre uma maior procura pelo atendimento médico do serviço de emergência do HR, independentemente da categoria enquadrada do paciente, caracterizando-se como o horário de pico. Às 9 horas a demanda é de 27 pacientes, enquanto que às 17 horas a

demanda é de 28 pacientes. No gráfico, os picos de atendimento são de 28 pacientes atendidos às 11 horas, às 14 horas e às 17 horas.

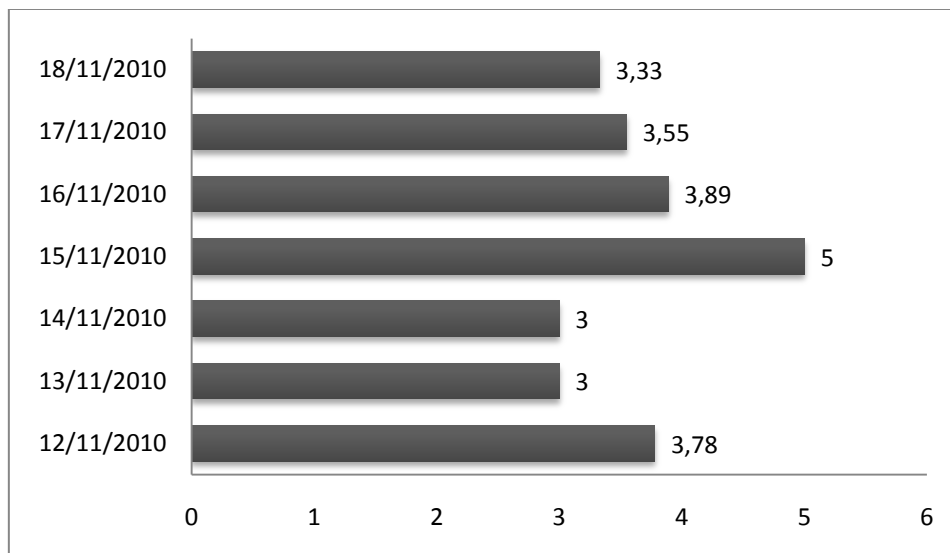


Figura 4.3 – Média da taxa de chegada/hora nos dias em horário de pico

Fonte: Sistema de Informação do HR (2010)

De acordo com os dados apresentados na Figura 4.3 percebe-se que na segunda-feira, dia 15 de novembro 2010, a taxa de atendimento é maior quando comparada com os outros dias da semana, e igual a cinco pacientes por hora. Nos dias 14 e 13 de novembro 2010 as médias das taxas de chegada são iguais e menores quando comparadas com a média dos outros dias, representando três pacientes/hora no horário de pico. A média da taxa de chegada no período foi de 3,65 pacientes por hora.

Desejou-se obter os dados relativos ao início do atendimento, de cada paciente considerado no período, e confrontar com o tempo máximo previsto do modelo de triagem de Manchester, mas não foi possível, pois, não se conseguiu chegar a esses dados através do sistemas de informação do HR, tais dados não foram disponibilizados.

4.4. Modelo de filas no HR

O modelo de filas no HR se caracteriza por ser o modelo de filas de disciplina de prioridades, de acordo com o Sistema de Triagem de Manchester.

Os dados foram coletados do dia 12 ao dia 18 de novembro 2010, no período de 24 horas e posteriormente foi identificada a taxa de chegada de pacientes por hora. Ao final foi identificada a taxa de chegada por cada classe de prioridades através da porcentagem dos pacientes classificados em cada uma das cinco categorias, no horário de maior procura pelo atendimento médico, o horário de pico, visto que nos períodos de menor demanda poderia ocorrer ociosidade. Notou-se a presença de dois médicos realizando os atendimentos.

A taxa média de chegada foi de 3,65 pacientes/hora, enquanto que para o tempo de atendimento, foram calculados dois valores, $\mu = 2,3$ pacientes/hora e $\mu = 3$ pacientes/hora.

Tabela 4.2 – Porcentagem do total de pacientes e taxa de chegada das cinco categorias de Manchester

Fonte: Sistemas de Informação do HR (2010) e Autor (2010)

Categoria	Porcentagem do total	Taxa de chegada por cada categoria (λ_i)
Emergência	8%	$\lambda_1 = 0,292$
Muito urgente	18%	$\lambda_2 = 0,657$
Urgente	64%	$\lambda_3 = 2,336$
Pouco urgente	9%	$\lambda_4 = 0,3285$
Não urgente	1%	$\lambda_5 = 0,0365$

De posse destes parâmetros (λ , λ_1 , λ_2 , λ_3 , λ_4 , λ_5 e μ) foi proposto o modelo de disciplina de prioridades não-preemptivas, sem restrição da capacidade e os clientes sendo atendidos por ordem de chegada dentro de uma mesma categoria.

Foi aplicado o modelo para $s = 2$ ($\mu = 2,3$) e $s = 2$ ($\mu = 3$) e chegou-se aos seguintes resultados (os valores para W e W_q estão em horas):

Categoria/parâmetro	λ_i	L	L_q	W	W_q
Emergência	0,292	0,174546	0,0475895	0,5977603	0,1629777
Muito urgente	0,657	0,4205608	0,1349086	0,6401229	0,2053403
Urgente	2,336	2,5870894	1,5714372	1,1074869	0,6727043
Pouco urgente	0,3285	0,9606773	0,8178512	2,9244363	2,4896536
Não urgente	0,0365	0,141656	0,1257865	3,8809874	3,4462048
Taxa de utilização (ρ) = 0,7934783.					

Tabela 4.3 – Resultados do modelo de disciplina de prioridades para $s = 2$ ($\mu = 2,3$)

Fonte: O autor (2010)

De acordo com os dados apresentados na Tabela 4.3, um paciente classificado na categoria *Muito urgente*, esperaria em média 12,32 minutos ($0,2053403 \times 60$) para começar o atendimento. 2,32 minutos a mais do que o esperado para esta mesma categoria do modelo de Manchester e 2,68 minutos a menos caso o modelo proposto fosse o Canadense. Um paciente classificado na categoria *Urgente* esperaria 40,36 minutos ($0,6727043$ horas) para ser atendido. 19,64 minutos a menos quando comparado com o tempo previsto máximo dessa mesma categoria no modelo de Manchester e 10,36 minutos a mais caso o modelo considerado fosse o Canadense ou o Australiano. Já na categoria *Pouco urgente*, um paciente esperaria em média 149,37 minutos para o início do atendimento, ou seja, 29,38 minutos a mais previsto pelo modelo de Manchester e 89,38 minutos a mais previsto pelo modelo Canadense ou Australiano. Para um paciente classificado na última categoria, este aguardaria 206,77 minutos ($3,4462048$ horas) até o início do atendimento médico. 33,23 minutos a menos quando comparado pelo tempo previsto máximo (240 minutos) nessa mesma categoria do modelo de Manchester e 86,77 minutos a mais caso o modelo considerado fosse o Canadense ou o Australiano.

A taxa de utilização do sistema ficou aproximadamente igual a 79,35%.

Tabela 4.4 – Resultados do modelo de disciplina de prioridades para $s = 2(\mu=3)$

Fonte: O autor (2010)

Categoria/parâmetro	λ_i	L	L_q	W	W_q
Emergência	0,292	0,1208749	0,0235416	0,4139553	0,0806219
Muito urgente	0,657	0,2819205	0,0629205	0,4291028	0,0957695
Urgente	2,336	1,249009	0,4703423	0,5346785	0,2013452
Pouco urgente	0,3285	0,2494885	0,1399885	0,7594781	0,4261447
Não urgente	0,0365	0,0301368	0,0179702	0,8256665	0,4923332
Taxa de utilização (ρ) = 0,6083333.					

De acordo com os dados apresentados na Tabela 4.4, percebe-se uma diferença significativa dos parâmetros quando comparados com os mesmos parâmetros da tabela 4.3 quando se considerou $\mu = 2,3$. Neste caso, ao considerar $\mu = 3$, os valores de L , L_q , W e W_q reduziram-se substancialmente. Por exemplo, um paciente classificado na categoria *Urgente*, ao invés de esperar, em média, 40,36 minutos para o início do atendimento, aguardaria 12,08 minutos na fila. Enquanto que um paciente classificado como Não urgente, ao invés de esperar 232,86 minutos no sistema (espera na fila mais o tempo de ser atendido), aguardaria 49,53 minutos, desde sua chegada até sua saída no sistema.

A taxa de utilização do sistema foi de 60,83%.

Foram realizados também os cálculos dos parâmetros (L , L_q , W e W_q) caso houvesse somente um médico realizando o atendimento. Para isso, se fez necessário que a taxa de atendimento fosse igual a quatro pacientes/hora.

Tabela 4.5 – Resultados do modelo de disciplina de prioridades para $s = 1(\mu=4)$

Fonte: O autor (2010)

Categoria/parâmetro	λ_i	L	L_q	W	W_q
Emergência	0,292	0,1448581	0,0718581	0,4960895	0,2460895
Muito urgente	0,657	0,3762209	0,2119709	0,5726346	0,3226346
Urgente	2,336	4,492566	3,908566	1,9231875	1,6731875
Pouco urgente	0,3285	4,4209556	4,3388306	13,458008	13,208008
Não urgente	0,0365	0,9939707	0,9848457	27,232074	26,982074
Taxa de utilização (ρ) = 0,9125					

De acordo com os dados referentes à Tabela 4.5, caso houvesse apenas um médico realizando o atendimento, os parâmetros L , L_q , W e W_q aumentariam substancialmente. Por exemplo, um paciente classificado na categoria *Não urgente* aguardaria 26,98 horas para ser realizado o início do atendimento e 27,23 horas para deixar o sistema, tempos demasiadamente longos. Caso um paciente fosse classificado na categoria *Muito urgente* ou *Urgente*, este esperaria 19,35 minutos e 100,39 minutos até o início do atendimento, respectivamente.

Analisando os resultados das Tabelas 4.3, 4.4 e 4.5, percebe-se que quando $s = 1$, o atendimento é priorizado para as três primeiras categorias em prol das duas últimas categorias. Um médico apenas para realizar o atendimento, traria um grande desconforto para os pacientes classificados nas categorias *Pouco urgente* e *Não urgente*, fazendo com que estes pacientes tivessem de passar tempos extremamente longos pra serem atendidos.

A taxa de utilização do sistema variou de 79,35% para 60,83% quando μ variou de 2,3 para três, uma diferença de 18,52. Entretanto, quando s variou de dois para um com uma taxa de atendimento de quatro pacientes/hora, a taxa de utilização do sistema ficou igual a 91,25%.

Através da análise das Tabelas 4.3, 4.4 e 4.5, percebeu-se que caso o modelo de priorização fosse o Canadense ou o Australiano, constantemente os pacientes

aguardando no sistema de filas, extrapolariam o tempo máximo previsto. Entretanto, o modelo de Manchester utiliza-se de tempos previstos maiores, diminuindo esta parcela de pacientes que aguardam mais do que o previsto.

4.5. Análise dos Resultados

Quando comparado com o modelo de priorização Canadense ou o modelo de priorização Australiano ou mesmo o modelo de priorização Americano, o modelo de priorização de Manchester torna-se mais eficaz para o serviço de urgência e emergência do HR, devido aos seguintes fatores:

- Caracteriza-se nas situações em que a demanda ocasionada pelo número de pacientes à espera do atendimento médico excede a capacidade de atendimento. O serviço de urgência e emergência do HR sempre teve uma demanda muito alta, porém, devido a criação das UPAs, esta demanda vem diminuindo. Pacientes classificados nas categoriais *Pouco urgente* e *Não urgente* são, muitas vezes, encaminhados às UPAs. De toda forma, é mais eficaz estar preparado com um modelo de priorização que leve em considerações situações adversas.
- O atendimento médico, realizado pelo clínico geral, é em função do critério clínico e não do critério administrativo, como ocorre no modelo Americano. O paciente não é, desta forma, priorizado a partir dos recursos necessários para o seu atendimento.
- Permite comparar dados entre os diversos hospitais que utilizam este mesmo modelo.
- Único modelo de priorização que utiliza sistemas de informação . Ferramenta extremamente importante que traz inúmeros benefícios com o manuseio de dados.
- Baseia-se em algoritmos clínicos, diferentemente dos demais com excessão do modelo Americano, desta forma, permiti-se que haja padronização.
- Baseia-se em categorias de sintomas e não em escalas de urgência pré-definidas que podem induzir a diagnósticos.
- Possui um tempo de espera previsto máximo maiores para as duas últimas categorias, e um tempo previsto máximo menor para a segunda e terceira categoria, quando comparado com os outros modelos. Dessa forma, o

atendimento médico prioriza pacientes que realmente necessitam de rapidez no atendimento.

5. Conclusão

Pacientes atendidos dentro de tempos previstos refletem uma boa gestão do serviço e conseqüentemente, um bom dimensionamento da capacidade do sistema. Já quando se tem o contrário, ou seja, pacientes que são atendidos fora dos tempos previamente definidos, refletem um mau dimensionamento da capacidade do sistema e insatisfações por parte dos pacientes, tornando o serviço de má qualidade.

A construção deste trabalho de conclusão de curso permitiu um conhecimento mais aprofundado sobre as variáveis que influenciam um sistema de filas e como um serviço de emergência de um hospital pode se antecipar dimensionando com mais exatidão sua capacidade de atendimento, a fim de evitar que longas filas se formem e prejudique a qualidade do serviço prestado.

Espera-se que este trabalho possa ajudar os diversos serviços de emergência dos hospitais públicos a definirem mais precisamente sua capacidade de atendimento e melhorar a qualidade do atendimento, de uma maneira tal que não tenha que, necessariamente, investir em infraestrutura; assim também como a definição da melhor escolha do modelo de priorização de atendimento que mais se identifica com a realidade e particularidades de cada hospital.

Ao apresentar argumentos teóricos sobre modelos de priorização do atendimento e teoria das filas, um estudo de caso foi proposto ao final desta monografia, tornando este trabalho mais consistente para com seu propósito, facilitando a leitura e entendimento por parte do leitor acerca do tema abordado.

Algumas dificuldades foram encontradas ao longo do desenvolvimento deste trabalho. Podendo-se citar o estudo da Teoria das Filas pelo autor. Apesar de ser ofertada, eletivamente, à grade curricular do curso de Engenharia de Produção da UFPE, não foi cursada pelo autor.

Por fim, ao desenvolver este trabalho de conclusão de curso, o autor encontrou algumas limitações por não relatar a diferença entre a comparação dos tempos devidamente cronometrados de permanência da fila e no sistema com estes mesmos tempos, porém, calculados através dos modelos de teoria das filas. Uma outra limitação

foi a respeito de uma possível alteração na porcentagem dos pacientes classificados em cada categoria caso o modelo de priorização fosse o Canadense, o Australiano ou o Americano, diferentemente do Manchester, usado atualmente na emergência do HR.

Esta alteração poderia resultar em parâmetros (L , L_q , W e W_q) diferentes dos que foram encontrados através da modelagem da teoria das filas no estudo de caso deste trabalho, conseqüentemente, a diferença entre o tempo previsto pela teoria das filas e o tempo previsto máximo pelo modelo de priorização alteraria-se.

Além disso, outras filas geradas no Hospital da Restauração poderiam ser objeto de estudo. Desta forma, recomenda-se o estudo de diferentes filas que se formam todos os dias ao longo de todos os setores do hospital.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, A. T.; ALMEIDA, L. M. A. B. Métodos multicritérios de apoio à decisão em saúde. In: Aplicações com métodos multicritério de apoio à decisão. 1ed. Recife: Editora Universitária. 2003.

AUSTRALIAN COLLEGE OF EMERGENCY MEDICINE. **Guidelines for implementation of the Australian triage scale in emergency departments.** [S.l.], 2001. Disponível em: <http://www.acem.org.au/media/policies_and_guidelines/G24_Implementation__ATS.pdf>. Acesso em: 29 set. 2010.

BITTENCOURT, R. J.; HORTALE, V. A. A qualidade nos serviços de emergência de hospitais públicos e algumas considerações sobre a conjuntura recente no município do Rio de Janeiro. **Ciênc. saúde coletiva**, Rio de Janeiro, v. 12, n. 4, Agosto. 2007. Disponível em <http://www.scielo.org/scielo.php?pid=S1413-81232007000400014&script=sci_arttext&tlng=pt>. Acesso em 6 out. 2010.

BRASIL. Ministério da Saúde. Sistema Único de Saúde. Cartilha Entendendo o SUS. Brasília: MS. Disponível em: <http://portal.saude.gov.br/portal/arquivos/pdf/cartilha_entendendo_o_sus_2007.pdf>. Acesso em: 5 set. 2010.

CORDEIRO JUNIOR, W. A classificação de risco como linguagem da rede de urgência e emergência. **RAHIS. Revista de Administração Hospitalar e Inovação em Saúde**, n. 2, jan/jun 2009. Disponível em: <<http://www.face.ufmg.br/revista/index.php/rahis/article/viewFile/787/663>>. Acesso em: 2 set. 2010.

COSTA, L. C. Teoria das Filas. Universidade Federal do Maranhão – UFMA. 2005. Disponível em: <http://www.deinf.ufma.br/~mario/grad/filas/TeoriaFilas_Cajado.pdf> Acesso em: 5 out. 2010.

ESTEVES, F. O. **Sistemas de informação a serviço da saúde Pública: a implantação do sistema de Urgência e emergência e a concepção do Repositório eletrônico de saúde (RES).** In: Congresso Consad de Gestão Pública, III, 2010, Brasília. Painel 24/095: Tecnologia da informação e de comunicação aplicada ao Sistema Único de Saúde de Minas Gerais: governança e inovações. Disponível em: <<http://www.consad.org.br/sites/1500/1504/00001910.pdf>>. Acesso em 1 out. 2010.

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

GILBOY, N. et al. **Emergency Severity Index, Version 4: Implementation Handbook**: AHRQ Publication No. 05- 0046-2. Rockville, MD: Agency for Healthcare Research and Quality, 2005. Disponível em: <<http://www.ahrq.gov/research/esi/esihandbk.pdf>>. Acesso em 2 out. 2010.

GOMES, C. F. P. **Contacto com o enfermeiro da triagem do serviço de urgência: Satisfação do utente**. Trabalho de Conclusão de Curso (Licenciatura em Enfermagem) – Universidade Fernando Pessoa, Ponte Lima, 2008. Disponível em: <<https://bdigital.ufp.pt/dspace/bitstream/10284/1084/3/CristinaGomes.pdf>>. Acesso em: 2 out. 2010.

GÓMEZ JIMÉNEZ, J. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergências. **Emergências**; n. 15, p. 165-174. 2003.

GOUVÊA, C. D.; TRAVASSOS, C.; FERNANDES, C. Produção de serviços e qualidade da assistência hospitalar no Estado do Rio de Janeiro, Brasil - 1992 a 1995. **Rev. Saúde Pública**, São Paulo, v. 31, n. 6, Dezembro. 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S003489101997000700009&lng=en&nrm=iso>. Acesso em: 19 set. 2010.

HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à Pesquisa operacional**, 8. ed. São Paulo: Mc Graw-Hill, 2006.

MACKWAY-JONES, K.; MARSDEN, J.; WINDLE, J. **Emergency triage group: Manchester Triage Group**. 2 ed. Oxford: BMJ publishing group, 2006.

MARINHO, A. **Estado de uma Nação: Texto de Apoio. A Economia das Filas no Sistema Único de Saúde Brasileiro**. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 2009.

MOREIRA, C. T. P. **Avaliação de uma implementação do Sistema de Triagem de Manchester: Que realidade?** 2010. 95 f. Dissertação (Mestrado em Informática Médica) - Faculdade de Medicina da Universidade do Porto, Porto, 2010. Disponível em: <<http://repositorio-aberto.up.pt/bitstream/10216/24572/4/teseMIMcatiaM.pdf>>. Acesso em: 1 out. 2010.

PIRES, P. S. **Tradução para o português e validação de instrumento para triagem de pacientes em serviço de emergência: Canadian Triage and Acuity Scale (CTAS)**. 2003. 206 f. Tese (Doutorado) – Escola de Enfermagem da Universidade de São Paulo, São Paulo, 2003. Disponível em: <www.teses.usp.br/teses/disponiveis/7/7136/tde.../Patricia_Silva_Pires.pdf>. Acesso em: 26 set. 2010.

RATI, R. M. S. **“Criança não pode esperar”: A demanda de mães e suas crianças em condições não urgentes em ambulatório pediátrico de urgência e emergência**. 2009. 121 f. Dissertação (Mestrado em Ciências da Saúde) - Faculdade de Medicina da Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <http://www.medicina.ufmg.br/cpg/programas/saude_crianca/teses_dissert/2009_mestra_do_RoseMeireRati.pdf>. Acesso em: 15 set. 2010.

SILVA, A. M. Triagem de prioridades - Triagem de Manchester. 2009. 232 f. Dissertação (Mestrado em Medicina de Catástrofe) – Universidade do Porto, Porto, 2009.

SOUZA, C. C. **Grau de concordância da classificação de risco de usuários atendidos em um pronto-socorro utilizando dois diferentes protocolos**. 2009. 121 f. Dissertação (Mestrado em Enfermagem) - Escola de Enfermagem da Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <<http://www.enf.ufmg.br/mestrado/dissertacoes/CristianeCSouza.pdf>>. Acesso em: 15 set. 2010.

TOLEDO, A. **Acurácia de enfermeiros na classificação de risco em unidade de pronto socorro de um hospital municipal de belo horizonte**. 2009. 140 f. Dissertação (Mestrado em Enfermagem) - Escola de Enfermagem da Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <<http://www.enfermagem.ufmg.br/mestrado/dissertacoes/AlexandreDToledo.pdf>>. Acesso em: 29 set. 2010.