

## **TOC&DES: UMA METODOLOGIA DE MELHORIA CONTÍNUA INTEGRANDO TEORIA DAS RESTRIÇÕES E SIMULAÇÃO DE EVENTOS DISCRETOS**

*TOC & DES: A METHODOLOGY OF CONTINUOUS IMPROVEMENT INTEGRATING  
RESTRICTION THEORY AND SIMULATION OF DISCRETE EVENTS*

Francisco Santos **SABBADINI**

Universidade do Estado do Rio de Janeiro / Universidade Estácio de Sá  
franciscosabbadibi@gmail.com

Augusto **GONÇALVES**

Universidade Estácio de Sá  
augusto@inca.gov.br

Mario Jorge Ferreira de **OLIVEIRA**

COPPE – UFRJ  
mario\_jo@po.coppe.ufrj.br

Artigo recebido em 01/2015 – Aprovado em 04/2016

### **Resumo**

Existem poucos trabalhos sobre a utilização integrada da Teoria das Restrições (TOC) e da Simulação de Eventos Discretos (DES) particularmente na área de saúde. Este artigo apresenta uma metodologia de melhoria contínua utilizando de forma integrada os dois campos do conhecimento acima. São apresentados estudos de caso no ambiente hospitalar com detalhamento da aplicação da metodologia desenvolvida. Em todos os casos os ganhos obtidos possuem implicações importantes como a satisfação do paciente no processo de tratamento e sua percepção em relação à qualidade dos serviços prestados. Outro aspecto a ser considerado é a melhoria da assistência prestada a estes que passam a ter seu tratamento iniciado mais rapidamente, contribuindo para a elevar a sua qualidade de vida, chances de cura e restabelecimento. Os resultados obtidos demonstraram a viabilidade da metodologia nos três casos apresentados.

**Palavras-chave:** Teoria das Restrições, Simulação de Eventos Discretos, Gestão da Capacidade, Melhoria Contínua.

<http://periodicos.unifacef.com.br/index.php/rea>

## ABSTRACT

*There are few studies about the integrated use of Restriction Theory (TOC) and Discrete Event Simulation (DES), particularly in the health area. This article presents a continuous improvement methodology using the two fields of knowledge mentioned above. Case studies are presented in the hospital environment with details of the application of the developed methodology. In all cases the gains obtained have important implications such as patient satisfaction in the treatment process and their perception regarding the quality of the services provided. Another aspect to be considered is the improvement of health care services contributing to raise life quality, chances of cure and recovery. The results obtained show the feasibility of the methodology in the three cases presented.*

**KEYWORDS:** *Theory of Restrictions, Simulation of Discrete Events, Capacity Management, Continuous Improvement.*

## INTRODUÇÃO

Existem poucos trabalhos sobre a utilização integrada da Teoria das restrições e da simulação, orientadas para gestão da capacidade de atendimento na área de saúde. Trabalhos nesse sentido foram desenvolvidos por Gonçalves (2004), Sabbadini (2005) e Sabbadini et. Al. (2008). Rébula et. al. (2009) utilizaram a teoria das restrições integradas a outras técnicas de produção como mapeamento de processos e gestão da capacidade. Este artigo apresenta um conjunto de casos aplicados, resultado de alguns anos de estudo que consolidam e mostram a aplicação de forma integrada da simulação e da TOC na área de saúde, a saber: em uma unidade de pronto atendimento, num centro de tratamento de câncer do Instituto Nacional do Câncer (INCA) e numa unidade de emergência hospitalar no Hospital Municipal de Emergência Henrique Sérgio Gregori. Assim sendo, está estruturado além desta introdução em três partes: i) fundamentação teórica sobre teoria das restrições; ii) simulação; iii) gestão da capacidade; iv) estudos de caso; e por fim a v) conclusão.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. TEORIA DAS RESTRIÇÕES

A teoria das restrições foi originalmente desenvolvida por Eliyahu Goldratt, um físico israelense que publicou os fundamentos no seu livro A Meta. Goldratt introduziu o termo teoria das restrições (“Theory of Constraints” – TOC) em 1987 englobando vários conceitos como o sistema de gerenciamento de desempenho, os cinco passos para melhoria contínua, refinamentos de conceitos de como gerenciar a produção e outras áreas de uma empresa.

Existem duas premissas em sua abordagem. A primeira é considerar a organização como um sistema cujo sucesso ou fracasso depende da forma como diferentes processos interagem entre si. A segunda é a de que uma restrição é qualquer coisa que limite o sistema de atingir sua meta. Sendo assim, todo sistema deve ter pelo menos uma restrição ou fator limitante. Uma restrição não é boa nem má. Ela existe e deve

ser gerenciada (GOLDRATT, 1990; 1994). De fato, o reconhecimento da existência de um fator limitante é uma excelente oportunidade para a melhoria, pois permite o foco na identificação e gerenciamento das restrições.

Apesar da teoria das restrições ter nascido no setor industrial, sua filosofia pode ser aplicada em diferentes tipos de organização, já existindo aplicações da TOC na área de saúde. Segundo Burton T. (2001), consultora do instituto Goldratt, as metas de uma organização de saúde de oferecer tratamento com qualidade e a de gerar lucro são complementares e essenciais. Na Inglaterra esta teoria foi aplicada com sucesso na redução das longas listas de espera que são administradas pelo sistema nacional de saúde conforme descrito por Phipps (1999). Na área do tratamento de câncer, Kershaw (2000) apresenta um estudo interessante realizado no setor de quimioterapia de uma clínica oncológica. Um dos princípios fundamentais da teoria das restrições baseia-se no processo de focalização buscando a melhoria contínua. Os cinco passos para se aplicar a teoria das restrições na área de saúde podem ser descritos da seguinte forma:

1. Identificar as restrições do sistema

Restrições físicas devem ser imediatamente identificadas como, por exemplo, o número de salas de exames, equipamentos, médicos, enfermeiras e técnicos. Em muitas instituições de saúde, a restrição está no número de auxiliares administrativos. Quando existe a necessidade de redução de custos, grande parte das organizações decidem por reduzir o pessoal administrativo preservando o corpo clínico. Uma ação desta natureza, sem uma análise detalhada do processo atual pode resultar em um crescimento de atividades administrativas, sendo desempenhadas por médicos e enfermeiras e a diminuição do tempo destinado ao atendimento do paciente. A velocidade do fluxo de pacientes atendidos cai, o faturamento cai e o custo operacional por paciente aumenta. Restrições políticas como decisões e estratégias ineficientes, apesar de muitas vezes mais desastrosas do que problemas operacionais são mais difíceis de identificar e gerenciar.

2. Decidir como explorar as restrições do sistema

Se a restrição é física, o objetivo é gerenciar os fatores limitantes tornando o processo o mais eficiente possível. O elo mais fraco corresponde à força e resistência de toda a corrente. A restrição dita a velocidade do fluxo de atendimento dos pacientes. O tempo perdido com um fator limitante é um tempo perdido em todo o processo.

Sendo assim, uma vez identificada a restrição ela deve ser imediatamente explorada e gerenciada. Isto pode ser realizado de diferentes maneiras, dependendo do tipo de restrição. Se a limitação está no número de equipamentos de diagnóstico a decisão está em como programar sua utilização para que o maior número de pacientes possam utilizá-lo sem que haja o comprometimento da qualidade do exame e do diagnóstico.

Todos os esforços devem ser direcionados para aumentar a taxa de utilização dos equipamentos considerados recursos gargalos. Ações de curto prazo, que aumentem a eficiência do agendamento de pacientes e reduzam o tempo do exame, devem ser exploradas e executadas.

3. Subordinar/Sincronizar todos os recursos restantes às decisões tomadas acima

Subordinar e sincronizar todo o restante para as decisões tomadas nos dois primeiros passos é a etapa mais difícil e geralmente a mais importante. Sem esta subordinação, os planos de explorar a restrição podem não ocorrer, sendo colocados de lado em função dos problemas e incêndios do dia-a-dia.

4. Elevar a capacidade das restrições do sistema

Em contraste com o passo 2 em que são realizadas ações para aumentar o fluxo de atendimento aos pacientes sem que hajam gastos significativos, este passo requer um investimento nos recursos considerados gargalo. Por exemplo, pode ser necessário o aumento da equipe médica ou o crescimento da capacidade instalada com a compra de um novo equipamento.

5. Se a restrição se deslocar, retornar ao passo 1 e não permitir que a inércia se instale.

Não se pode permitir que a inércia se instale e se transforme numa restrição do sistema. O ambiente está em constante mudança. Uma solução tende a se deteriorar ao longo do tempo. A teoria das restrições é um processo de aprimoramento contínuo.

Quando se analisa o fluxo de tratamento de pacientes em um hospital ou clínica, pode-se observar que o mesmo é composto de uma sequência de atividades tais como, registro, consultas ambulatoriais, exames de diagnóstico e condutas terapêuticas, que configuram uma sequência linear de eventos ou uma cadeia de processos interdependentes. Cada elo dessa cadeia possui a habilidade de executar suas respectivas atividades em diferentes taxas médias de atendimento. É importante ressaltar que esta cadeia é tão forte quanto o seu elo mais fraco que corresponde ao processo com a menor taxa média de pacientes atendidos.

## 2.2. SIMULAÇÃO

A crescente elevação da complexidade da gestão dos serviços de saúde, associada à redução da oferta de recursos nesta área revelou-se um campo potencial para a aplicação das técnicas de Pesquisa Operacional (PO). Diversos trabalhos foram desenvolvidos neste campo por De Oliveira e Filho (2003), De Oliveira e Toscano (2003), Chagas (2003), Gonçalves (2004), Sabbadini (2005) e Magalhães (2006), que apresentam ferramentas e iniciativas que proporcionaram respostas significativas em termos de resultados. O uso da simulação está amplamente documentado em Brailsford (2009).

Neste contexto seja por falta de informação, de recursos ou de organização dos processos administrativos e operacionais, ficam perceptíveis as dificuldades e os desafios com os quais se defrontam os administradores hospitalares. Nesse sentido a melhoria na eficiência destes serviços é de interesse particular para a maioria das comunidades, em qualquer lugar do mundo, e segundo De Oliveira e Toscano (2003) um componente crítico do sistema é a resposta resultante do gerenciamento dos serviços em termos de atendimento da demanda. Em função disso os gestores tem procurado por instrumentos confiáveis para compreensão, análise do sistema de

serviço e implementação de mudanças apropriadas. A simulação computacional é uma técnica que preenche essa lacuna.

Para Banks e Carson (1984), a simulação é a imitação da operação de processos reais ou sistemas num dado período de tempo. Essa técnica envolve a geração de uma história artificial de um sistema, e a observação desta para fazer inferências relativas às características de operação do sistema real. De acordo com Gonçalves (2004) o custo relativamente baixo, pouco risco em relação à experimentação diretamente no sistema real e a capacidade de suportar de maneira confiável as flutuações estatísticas são algumas das características que recomendam a utilização desta técnica.

### **2.3.- GESTÃO DA CAPACIDADE**

Capacidade é o nível máximo que um processo pode atingir ao operar, sob condições normais, em determinado período de tempo (SLACK, 2002). O intuito de gerenciar essa capacidade é atender a demanda de maneira eficiente.

Outra questão que deve ser considerada, tanto em operações de manufatura quanto nas de serviços, é que algumas partes trabalham abaixo da sua capacidade e outras em seu limite máximo. A esse respeito Slack (2002, p.345), registra que "(...) dependendo da natureza da demanda, o uso de diferentes partes de uma operação pode atingir sua capacidade máxima e atuar como uma restrição para toda a operação".

Além das questões relativas aos gargalos, particularmente em operações de serviço, há o desafio de se estabelecerem medidas de capacidade. Segundo Slack (2000), "o principal problema com a medição da capacidade é a complexidade da maior parte dos processos produtivos. Somente quando a produção é altamente padronizada e repetitiva é fácil definir a capacidade". As principais decisões sobre capacidade devem considerar os seguintes aspectos: Estimar as capacidades das instalações atuais, prever as necessidades futuras de capacidade para todos os produtos e serviços, identificar e analisar as diferentes fontes e formas de alterar a capacidade no curto, médio e longo prazo, identificar diferentes formas de influenciar a demanda, avaliar o impacto da decisão a respeito de capacidade sobre a qualidade do serviço prestado e, por fim, escolher fontes alternativas para incrementar a capacidade do sistema.

As decisões sobre a capacidade requerem a identificação das áreas do sistema em que os benefícios da alteração vão repercutir no aumento da capacidade global. Uma vez efetuada esta etapa, as políticas de alteração de capacidade devem considerar modificações de curto, de médio e longo prazo. As estratégias de gestão da capacidade e da demanda são instrumentos de que os gestores dispõem face ao desafio de equilibrar a oferta à demanda.

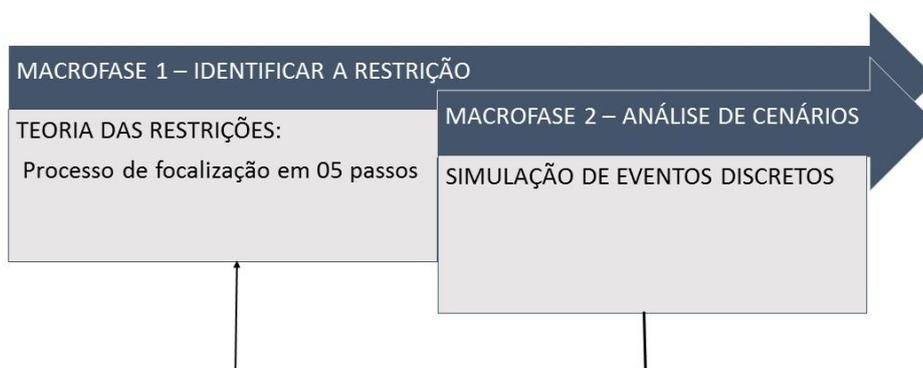
Neste contexto, faz-se necessária uma abordagem sistemática, estruturada e consistente que trate dessas questões e ofereça alternativas para o desenvolvimento de uma estratégia adequada de gestão da capacidade em cada caso, como será apresentado nos casos a seguir.

### 3. UMA METODOLOGIA CÍCLICA DE MELHORIA CONTÍNUA

Um dos aspectos relevantes relativo as operações de produção e serviços é que mesmo quanto suas atividades estão planejadas e controladas sempre há espaço para a melhoria das mesmas. Segundo Slack et al. (2009) uma vez que a prioridade do melhoramento tenha sido definida é necessário definir uma abordagem estratégica para a mesma. As duas estratégias apontadas são o melhoramento revolucionário e o melhoramento contínuo. A primeira baseada em inovação presume uma mudança grande e dramática na forma como a operação trabalha. A segunda, como o nome indica, adota uma abordagem de melhoramento de desempenho gradual, que supões mais e menores passos de melhoramento incremental.

Slack et all (2009) consideram que uma característica do conceito de melhoramento contínuo e a ideia de o mesmo pode ser representado por um processo, sem fim e repetitivo, se constituindo num ciclo de melhoria. Ainda, segundo os mesmos autores, existem muitos ciclos de melhoramento, como o ciclo PDCA e o ciclo DMAIC. Em ambos os casos, o processo de melhoria ocorre à medida que os passos são sequencialmente seguidos e desenvolvidos.

Neste sentido o presente trabalho adota uma abordagem de melhoria contínua de natureza cíclica de melhoria. O modelo é constituído de 02 macrofases. A primeira é destinada à identificação da restrição ou gargalo na operação e a segunda macrofase na identificação de oportunidades de melhoria via simulação do gargalo. A figura 1 apresenta esquematicamente a metodologia.



**Figura 1:** Metodologia TOC&DES de melhoria contínua

### 4. ESTUDOS DE CASO

Nesta seção serão apresentadas aplicação práticas na área de saúde com a implementação do processo de focalização da TOC integrado a simulação. O primeiro estudo de caso descreve o processo numa unidade de pronto atendimento. Em seguida em um centro de tratamento de câncer numa unidade assistencial do Instituto Nacional do Câncer (INCA), na cidade do Rio de Janeiro, RJ. O segundo estudo de caso

foi realizado na unidade de emergência hospitalar do Hospital Municipal de Emergência Henrique Sérgio Gregori (HMEHSG), em Resende, RJ.

#### **4.1. HOSPITAL GERAL DE NOVA IGUAÇU**

Neste caso especificamente utilizou-se a simulação de eventos discretos para identificar a restrição e posteriormente aplicou-se o processo de focalização. A metodologia de simulação utilizado está fundamenta em Bank e Carlson (1984) e em Rossetti et al (1999). O Hospital Geral de Nova Iguaçu (Hospital da Posse) foi inaugurado no segundo trimestre de 1982. A unidade possui 15 mil metros quadrados de área construída e é dividida em cinco blocos: Bloco A (ambulatório e laboratório), Bloco B (direção, raios-X e CPD), Bloco C (emergência), Bloco D (refeitório, cozinha, UTI, centro cirúrgico, RPA e esterilização) e Bloco E (enfermagem e prédio da subestação elétrica). Em área anexa, está o Centro de Referência Parteira Mariana Bulhões (CRMB). O Serviço de pronto atendimento apresentava uma média mensal de 16.279 atendimentos, com uma média diária de 543 pacientes e uma taxa horária média de 54 pacientes, o que têm refletido no congestionamento e longas filas de espera que causam desconforto aos pacientes e pressão sobre a equipe médica.

##### **4.1.1- Serviço de pronto atendimento (SPA)**

O Serviço de Pronto-Atendimento (SPA) oferece as especialidades de pediatria e clínica médica 24 horas por dia, além do setor de ortopedia e de pequenas cirurgias que funciona de 7h às 19h. Um aspecto que chamou a atenção no SPA é que este funciona como uma das portas de entrada para a emergência do hospital, com a qual possui um corredor de acesso. Além disso, neste setor efetua-se uma classificação de risco dos pacientes nas filas, de modo a priorizar o atendimento dos casos mais graves. Este procedimento é realizado por dois enfermeiros que efetuam uma breve entrevista com os pacientes de modo a estabelecer uma avaliação inicial. Os recursos disponíveis neste segmento são: 02 recepcionistas, 02 enfermeiras, 02 médicos no Pronto-Atendimento, 02 médicos na área de observação, 02 consultórios no pronto-atendimento e 01 sala de observação.

Outra questão identificada in loco durante a visita ao setor é a diversidade de casos em termos de tipo de ocorrência, que vão desde pacientes buscando atestado médico ou atenção, até casos de emergência. Uma emergência é um estado ou situação inesperada, de quebra da ordem estabelecida, podendo ser esta de vários tipos, dependendo do sistema no qual se apresente. Segundo Sabbadini (2005) esta classificação pode ser agrupada em três categorias:

**Emergenciais:** Casos que apresentam um estado que indica perigo de morte e necessidade de atenção imediata;

**Urgentes:** Aqueles que apesar de não apresentarem perigo de vida devem receber atenção médica especializada, no decorrer de poucas horas, sob o risco de tornarem-se uma emergência.

**Ambulatoriais:** Casos crônicos que não exigem atenção imediata, nem oferecem riscos imediatos ao estado de saúde do paciente.

#### **4.1.2- Simulação Computacional**

O SPA vem enfrentando uma elevação de demanda por seus serviços e os gestores identificaram problemas de sobrecarga no setor. Em face disso desenvolveu-se um estudo que resultou num modelo de simulação a eventos discretos. A metodologia utilizada baseou-se em Rossetti et al (1999) e considera os seguintes passos:

##### **4.1.2.1.-Definição do problema e dos objetivos**

A partir da análise do fluxo dos pacientes e da elaboração do modelo conceitual pode-se identificar a existência de um gargalo na unidade de observação de pacientes com casos urgentes que se relaciona diretamente com problemas de congestionamento e também resultam em elevado tempo de permanência do paciente no setor. A partir desta avaliação inicial definiu-se como objetivo identificar a configuração de recursos humanos e materiais que minimizassem os efeitos da restrição existente. A hipótese inicial foi a de que a disponibilidade de leitos representava o recurso restritivo.

##### **4.1.2.2 - Desenvolvimento do modelo conceitual**

Definido o problema e determinados os objetivos da simulação, formulou-se a partir da análise do fluxo de pacientes o modelo conceitual do SPA, sendo que este estudo teve como foco o processo de atendimento de pacientes com casos urgentes naquela unidade de serviço, o qual é descrito a seguir:

O paciente chega ao SPA e aguarda numa fila. Dois enfermeiros fazem entrevistas com estes para efetuar uma classificação de risco, quanto à severidade do quadro clínico de cada paciente. Estes podem ser classificados com um caso de emergência (risco de vida), um caso urgente ou um caso ambulatorial (crônico), tendo a respectiva prioridade no atendimento.

A partir da avaliação dos enfermeiros cada paciente segue um fluxo distinto dentro das instalações do SPA. Sendo um caso de emergência é encaminhado diretamente para o pronto-socorro do hospital para os procedimentos necessários e se for o caso internação, saindo posteriormente do sistema SPA e entrando noutro processo, pertinente a área de emergência. Sendo classificado como um caso urgente, o paciente é encaminhado a uma sala de observação, onde receberá os primeiros cuidados e será monitorado por um dos dois médicos alocados naquela área. Havendo agravamento do quadro, será encaminhado para a área de emergência. Estando reabilitado, o paciente sai do sistema.

Nos casos relacionados aos pacientes ambulatoriais (crônicos) verifica-se a disponibilidade da especialidade clínica no SPA sendo o paciente encaminhado para atendimento em uma das duas salas disponíveis. Não havendo a especialidade clínica no setor o paciente é encaminhado para atendimento na rede primária municipal de Nova Iguaçu e sai do sistema. O modelo conceitual descrito é serviu de base para a elaboração do modelo computacional.

#### **4.1.2.3- Coleta de dados**

Na primeira etapa do estudo as informações necessárias à elaboração do modelo conceitual foram obtidas in loco. Este trabalho foi realizado com o acompanhamento de uma enfermeira do SPA. Os dados relativos aos volumes de procedimentos e demais dados estatísticos foram obtidos nos registros disponíveis na página eletrônica e de outras fontes do HGNI. As taxas de chegada e os tempos de atendimento foram estimados com base na experiência dos profissionais de saúde envolvidos no projeto. Cabe ressaltar que foram identificadas distribuições de probabilidade que se ajustassem aos dados coletados. Os resultados obtidos indicaram que a chegada de pacientes na área de observação ocorre segundo uma exponencial de média 11. Para o tempo de atendimento pelos médicos utilizou-se uma distribuição triangular com valor mínimo de 10 minutos, com o valor mais provável de 20 minutos e o valor máximo de 30 minutos.

#### **4.1.2.4 - Codificação do modelo computacional**

A partir do fluxo de tratamento a pacientes e do modelo conceitual foi desenvolvido o modelo computacional, com a utilização do software Arena, versão de treinamento 10.0, cujos módulos pré-definidos facilitaram a sua construção.

A simulação é terminal, considerando um período equivalente a duas horas de aquecimento. Foram efetuadas cinquenta replicações, cada uma equivalendo a um dia de operação da unidade de pronto atendimento, com a finalidade de obter mais acuracidade e eliminar distorções decorrentes de flutuações estatísticas. O tempo definido no relógio de simulação inicia-se às 8 horas e termina às 16 horas.

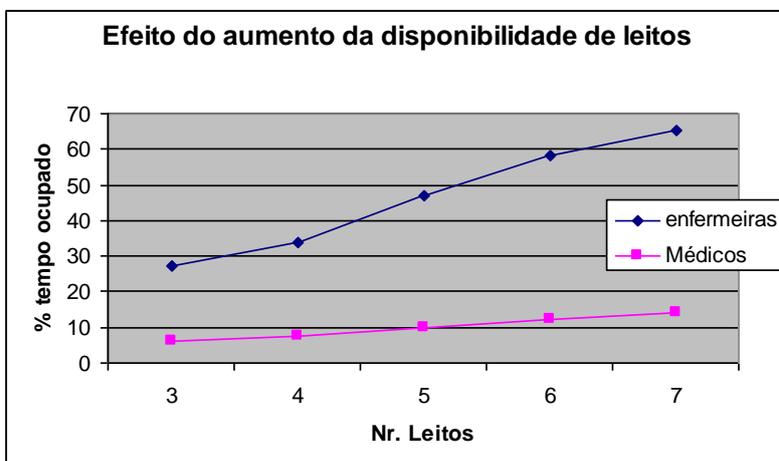
#### **4.1.2.5- Validação e verificação do modelo**

Nesta fase realizou-se uma apresentação estruturada do modelo desenvolvido para os gestores do hospital e a equipe envolvida no projeto, na qual foi descrita a forma como este se relaciona com o sistema real. A verificação é o processo de assegurar que o modelo de simulação foi construído corretamente e é completada quando o programa funciona da maneira como o analista de simulação pretendia. Ela pode ser efetuada rodando-se a simulação e monitorando-se sua operação. Se os parâmetros de entrada e a estrutura lógica do modelo estiverem corretamente representados no código, a verificação é finalizada. Para garantir a correta representação do modelo foi realizada a verificação lógica dos módulos com o fluxo de pacientes. O experimento foi realizado diversas vezes até que os resultados obtidos replicassem o modelo real.

#### **4.1.2.6 – Experimentação, execução da simulação e análise dos resultados**

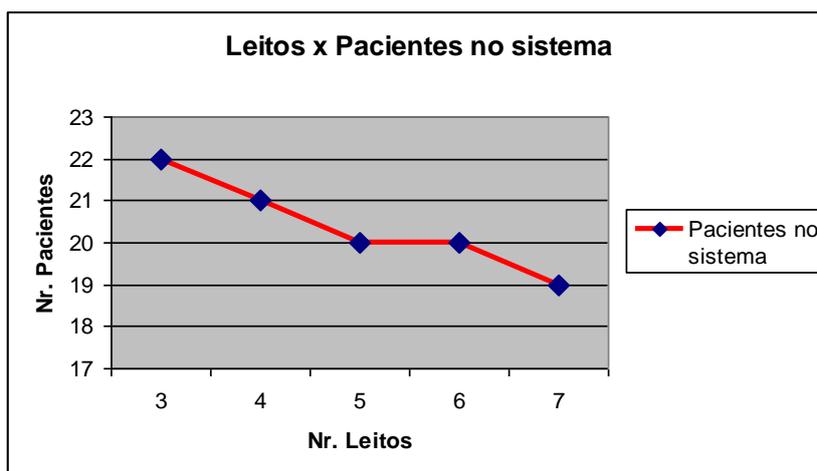
Considerando a configuração atual da área de observação de casos de pacientes urgentes foram realizados diversos experimentos com várias rodadas de simulação variando a quantidade de leitos disponíveis. Os experimentos realizados compreenderam um período de 8 horas iniciando-se às 8 horas e terminando às 16 horas. Esse corte de tempo é devido ao maior fluxo de pacientes estar concentrado neste período. Para cada experimento foram realizadas 50 replicações com o intuito de eliminar distorções e garantir a consistência dos resultados. Cada replicação equivale a um período completo de oito horas de operação nas condições

mencionadas. Dos resultados das simulações puderam-se extrair informações importantes para uma avaliação daquele setor. A figura 2 apresenta os resultados da primeira hipótese, onde se consideraram como fixos o número de médicos e de enfermeiras e variou-se o número de leitos disponíveis.



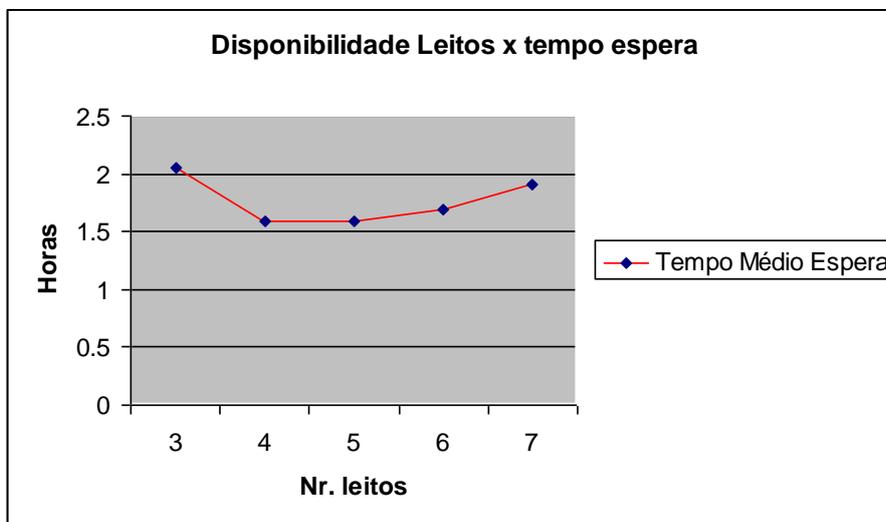
**Figura 2:** Disponibilidade de leitos e taxa de ocupação da equipe médica

A situação específica explorada nos experimentos partiu da hipótese de que a disponibilidade de leitos na área de observação era uma restrição à agilidade no atendimento e causava efeito sobre a taxa de ocupação da equipe médica. Pode-se verificar na figura 1 que à medida que se aumenta o número de leitos eleva-se a taxa de ocupação dos médicos e das enfermeiras. No mesmo sentido pode-se verificar um aumento da produtividade da equipe pela redução no número de pacientes no sistema ao final das rodadas de simulação, o que conforme se pode verificar na figura 3, indica um maior número de pacientes atendidos e liberados.



**Figura 3:** Efeito da disponibilidade de leitos sobre pacientes no sistema

Uma segunda situação foi verificada após a realização dos primeiro experimentos, a partir da observação do comportamento do tempo médio de espera do paciente por leito. Conforme se pode ver na figura 4, observa-se que a partir do quinto leito ocorre um ligeiro aumento no tempo de espera.



**Figura 4:** Relação entre número de leitos e tempo médio de espera do paciente

Partindo-se da hipótese que tal efeito é resultado do fato de o número de médicos e enfermeiras ter permanecido fixo, realizaram-se experimentos acrescentando-se um médico a partir do sexto leito para avaliar o efeito sobre o tempo médio de espera (em horas). Verificou-se que houve pouco efeito sobre o tempo de espera, e confirmou-se a disponibilidade de leitos como uma restrição física. Aplicou-se o processo de focalização da teoria das restrições, identificando-se a necessidade estratégica de implantar-se um sistema de classificação de risco acompanhado de agendamento para as consultas ambulatoriais.

#### 4.2. INSTITUTO NACIONAL DO CÂNCER

No tratamento do câncer, um dos fatores críticos de sucesso é o diagnóstico precoce. A agilidade no encaminhamento do paciente a adequadas condutas terapêuticas, com a redução do tempo de espera é fundamental para o aumento da sobrevivência, melhoria da qualidade de vida e chances de cura. O fluxo de tratamento de pacientes é composto de três grandes processos: Processo de Triagem: Identifica através de exames clínicos se a suspeita de câncer procede ou não; Processos de Diagnóstico: Detectam através de uma série de exames o tipo de câncer, sua localização e o estágio de evolução clínica da doença. Somente após estes exames o tratamento é realmente iniciado; Processos Terapêuticos: Encaminhamento do paciente a adequadas condutas terapêuticas, que correspondem, em sua grande maioria, a uma cirurgia, aplicações de radioterapia e quimioterapia.

Estes processos são interdependentes. O desempenho de cada um depende do anterior. O desempenho do sistema depende da estrutura desta cadeia de processos. O que importa é o sistema como um todo, e não os processos isolados.

O problema em questão está relacionado à identificação de possíveis gargalos no processo de diagnóstico e à melhor alocação dos equipamentos hospitalares, componentes estratégicos da capacidade instalada de um hospital. Estes gargalos são considerados como restrições à meta de oferecer um tratamento ágil e com qualidade. Neste contexto, pode-se utilizar a TOC que é um conjunto de princípios e ferramentas para resolução e minimização de gargalos existentes, para melhoria do desempenho do sistema como um todo. A seguir será descrita a aplicação do processo de focalização:

#### 1. Identificar as restrições do sistema

Para aplicar o primeiro passo do processo de focalização, que é a identificação da restrição, elaborou-se o fluxo de paciente, desde sua triagem até o início de um procedimento terapêutico. Quando um paciente é encaminhado ao hospital, a primeira etapa possui o objetivo de confirmar através de exames clínicos, o diagnóstico de câncer. Caso o diagnóstico não seja confirmado, o paciente é encaminhado para outros hospitais. Caso seja confirmado ou haja a necessidade de exames complementares, o paciente é matriculado para que se detecte a localização e o estágio de evolução clínica da doença. Após estes exames, o tratamento é realmente iniciado.

O fluxo do paciente foi analisado de forma detalhada e através do processo de focalização foi possível identificar que o departamento de radiologia desempenhava função estratégica, já que qualquer atraso na área afetava todo o processo. Os demais exames de patologia clínica e anatomia patológica apresentavam resultados com intervalos de tempo bem menores que os de imagem, não se constituindo em fatores limitantes no processo de diagnóstico.

Através da aplicação da TOC ficou clara a necessidade de aumentar a capacidade diária de execução de exames, com repercussão imediata na redução do tempo entre a marcação e sua realização. O gargalo estava localizado exatamente entre os processos de agendamento e realização dos exames de imagem, que levavam mais de 30 dias no caso da tomografia computadorizada. Neste contexto a redução do tempo era fundamental para agilizar o início do tratamento.

#### 2. Decidir como explorar a restrição do sistema

Identificada a restrição, o passo seguinte foi decidir como explorá-la. Todos os esforços foram direcionados para aumentar a taxa de utilização dos equipamentos considerados recursos gargalos. De modo a atender a todos os pacientes e evitar a ociosidade dos tomógrafos, a gestão de curto prazo é fundamental, para através de uma análise de sensibilidade antecipada, encontrar uma solução de equilíbrio. As ações de curto prazo consideradas para aumentar eficiência do sistema no sentido de reduzir o tempo dos exames foram à atribuição de prioridade de atendimento de acordo com o paciente e o tipo de exame, efetuar o agendamento buscando minimizar a ociosidade dos equipamentos disponíveis e ajustar os níveis de capacidade no caso de sobrecargas.

Para avaliar essas alternativas o problema em questão foi o foco de um estudo de simulação (GONCALVES E DE OLIVERA, 2004). Durante a modelagem e a fase de levantamento de dados, pontos importantes foram debatidos com as áreas envolvidas. A discussão e o entendimento de cada atividade foi tratada de forma integrada com o corpo clínico da área de radiologia, para que aspectos relevantes do processo replicado fossem representados no modelo construído.

Os resultados da simulação permitiram uma análise apurada de indicadores estratégicos, tais como taxa de utilização dos tomógrafos, quantidade de pacientes atendidos, tempo de permanência nas filas, fornecendo informações importantes para a tomada de decisão em cada cenário analisado.

### 3. Subordinar/Sincronizar todos os recursos restantes às decisões tomadas acima

Identificadas as alternativas mais adequadas o passo seguinte foi subordinar e sincronizar todos os recursos restantes às decisões tomadas acima. A utilização da simulação foi importante na identificação antecipada de eventuais problemas na agenda de um dia típico que foram solucionados através de ações gerenciais tais como: 1) a utilização de roteiros alternativos, que consistem na reprogramação do agendamento visando maximizar a utilização dos tomógrafos; 2) horas extras para atender a demanda no caso de estouro da capacidade planejada.

Para garantir a máxima utilização dos recursos gargalo foi desenvolvido e implementado o Sistema de Gestão de Atendimento (GONÇALVES E DE OLIVEIRA, 2003; 2004), que indica quando há necessidade de intervenção para manter o fluxo de atendimento planejado. Sua abordagem inovadora permite aos gestores acompanhar a situação das filas em cada uma das clínicas.

### 4. Elevar a capacidade da restrição do sistema

Ao se implementar os passos 2 e 3 obtém-se como resultado a elevação da capacidade da restrição do sistema pela redução da ociosidade no gargalo. Não houve a necessidade de elevar a restrição pela contratação de mais recursos humanos ou pela aquisição de equipamentos, uma vez que a solução encontrada atendeu os objetivos de um atendimento mais ágil.

### 5. Se a restrição se deslocar, retornar ao passo 1 e não permitir que a inércia se instale.

O simulador e o Sistema de Gestão de Atendimento permitem que o gestor tenha total domínio do dia-a-dia da clínica, aumentando sua eficiência nas tomadas de decisão relativas à programação da agenda. Desse modo é possível gerenciar a restrição e atuar de maneira efetiva no sentido de garantir a eficácia do fluxo desejado de serviço prestado ao paciente.

O resultado após a implementação das alternativas mostra que o intervalo de tempo médio entre a requisição do exame de tomografia computadorizada e sua realização caiu de 30,71 para 22,71 dias, com uma redução de 25% do tempo de espera, comprovando a eficácia do processo.

### 4.3. HOSPITAL MUNICIPAL DE EMERGÊNCIA HENRIQUE SÉRGIO GRÉGORI

Nos hospitais, o departamento de emergência é uma área estratégica, onde situações críticas frequentemente ocorrem. No atendimento de emergência o tempo é um fator crucial. A meta é a avaliação rápida, a estabilização, o tratamento e pronta admissão do paciente ao hospital. Quanto mais eficaz for esse processo, maiores serão as chances de restabelecimento, redução de sequelas e cura do paciente.

O Hospital Municipal de Emergência Henrique Sérgio Gregori (HMHESG) é um componente do sistema de saúde de Município de Resende, integrando o Sistema Único de Saúde (SUS), tendo como finalidade o atendimento integral a pacientes em estado de urgência ou emergência. Devido à eficácia na prestação de atendimento de casos de emergência, além da sua infraestrutura e localização geográfica, o HMEHSG tornou-se ao longo dos anos uma referência neste tipo de serviço para o município de Resende e a microrregião vizinha. É reconhecido também como hospital de referência no atendimento de emergências decorrentes de acidentes ocorridos da Rodovia Presidente Dutra.

A unidade de emergência do hospital funciona 24 horas, todos os dias, atendendo diariamente uma média de 192 pacientes. A direção do HMEHSG vem observando mudanças no perfil da população de pacientes atendidos e ao aumento da procura por seus serviços. Como consequência, sua capacidade de atendimento é inferior à demanda, verificando-se com frequência congestionamento na unidade de emergência, a formação de longas filas de espera e o aumento do tempo de permanência de pacientes no hospital, tendo como consequência pressão sobre os serviços e a equipe médica.

Outro fato constatado pela direção do hospital é que a maioria dos pacientes que procuram a unidade de emergência não necessita de atendimento imediato, prejudicando os pacientes de urgência e os de emergência. Neste contexto a teoria das restrições é adequada para a caracterização e tratamento do problema em questão, uma vez que o objetivo principal é agilizar o fluxo de tratamento de pacientes, identificando e gerenciando gargalos. O processo de focalização da TOC é descrito a seguir:

1. Identificar as restrições do sistema

Para esse fim, foi elaborado o fluxo de tratamento a pacientes no HMEHSG desde a triagem até o tratamento. Toda a elaboração foi realizada in loco com o acompanhamento do médico chefe da emergência e posteriormente validada por outros membros do corpo clínico, envolvendo médicos e enfermeiros da unidade. O fluxo foi estudado para cada perfil de paciente segundo critérios de emergência (risco de vida), urgência (caso grave) e ambulatorial. Um fluxo geral é apresentado na figura 5.

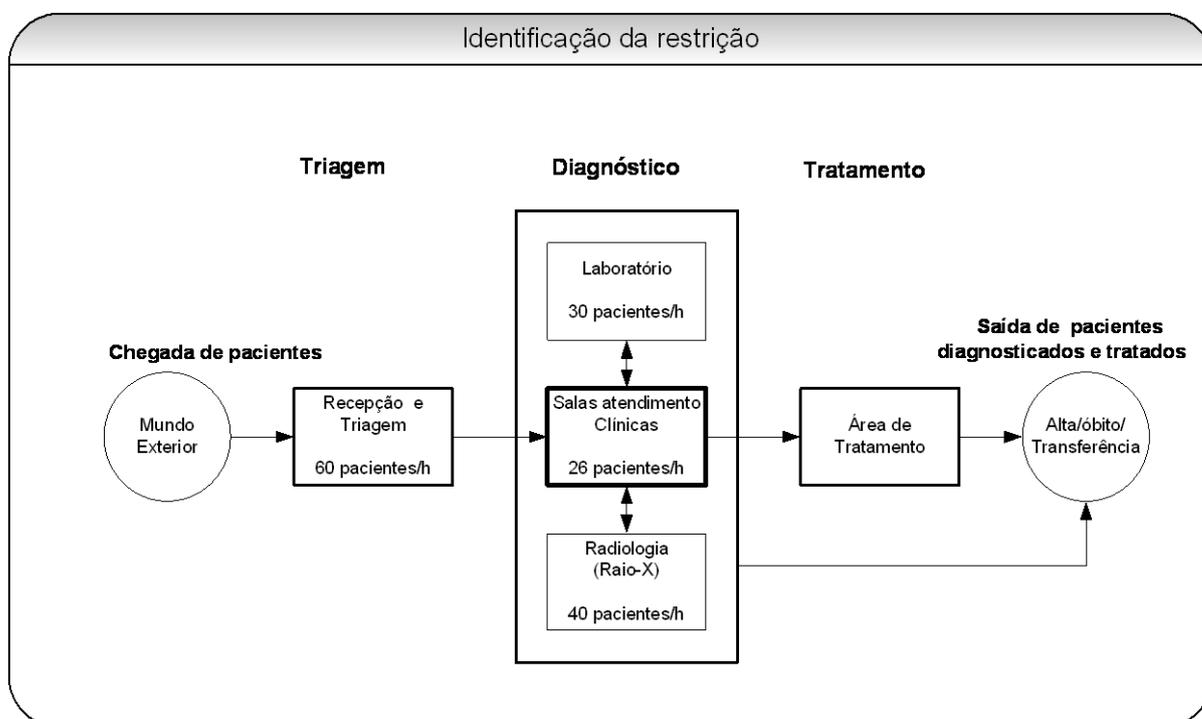
O paciente chega à unidade de emergência, sendo recebido por uma das recepcionistas, que efetua o preenchimento e o registro do boletim de atendimento médico. Em seguida solicita ao paciente que aguarde na recepção. O médico disponível recebe o paciente e inicia o atendimento. Faz a avaliação deste e emite o diagnóstico. Constatado ser um caso urgente é encaminhado para o pronto-socorro sendo atendido

por um médico da emergência. Tratando-se de um paciente ambulatorial após a consulta este é liberado.

O fluxo de pacientes, que é composto pelos processos de triagem, diagnóstico e tratamento foi analisado detalhadamente. São operações interdependentes, em que o desempenho de uma depende do resultado do imediatamente anterior. Na análise do fluxo de tratamento constatou-se o descompasso estrutural entre a oferta de recursos e a demanda, que resultam na formação de filas de espera, aumento do tempo de permanência de pacientes para o início do tratamento. A menor taxa de atendimento está nas clínicas, conforme apresentado na figura 5.

Associada a essa restrição há pressão sobre os serviços e a equipe médica, que divide sua atenção entre a demanda de pacientes que chegam para o primeiro atendimento e a daqueles que retornam dos serviços de laboratório e radiologia, com os resultados dos exames.

Conjuntamente, os pacientes ambulatoriais, que não necessitam de atendimento imediato, concorrem com casos urgentes pela atenção dos médicos, o que leva ao aumento do tempo de espera pelo primeiro atendimento, protelando o início do tratamento. Nesta situação o paciente com quadro urgente aguardava na fila de espera, sofrendo agravamento no seu quadro clínico, o qual somente seria detectado efetivamente quando este fosse atendido pelo médico.



**Figura 5:** – Fluxo de atividades na unidade de emergência no HMEHSG

O processo se configura como uma restrição à meta da unidade de emergência de oferecer um tratamento ágil ao paciente urgente. A agilidade no encaminhamento ao tratamento é fundamental para a redução de sequelas e melhoria da qualidade do serviço prestado a esse paciente.

## 2. Decidir como explorar a restrição do sistema

Identificada a restrição, no sentido de aproveitar ao máximo a capacidade existente no gargalo e manter o fluxo de atendimento desejado, optou-se por duas alternativas de reorganização dos processos. A primeira delas foi à implementação de uma triagem especializada para direcionar os pacientes com quadro urgente para áreas de tratamento sem restrição. A segunda alternativa consiste na alocação dos médicos nas clínicas em função da demanda por cada uma delas, principalmente nos horários de pico, que são os mais críticos.

Para avaliar a viabilidade dessas alternativas definiu-se um modelo de simulação cujo escopo compreendeu os setores de Pronto Atendimento e Pronto-Socorro da unidade de emergência, considerando as instalações, os recursos humanos e materiais relativos a cada processo analisado. Com base no fluxo de tratamento a pacientes e do modelo conceitual foi desenvolvido o modelo computacional, com a utilização do software MedModel, versão 3.5, cuja abordagem utiliza módulos pré-definidos que facilitam a sua construção. Durante todo o desenvolvimento houve o envolvimento dos médicos responsáveis pela unidade. A animação do modelo computacional desenvolvido permitiu o seu completo entendimento pelos gestores e demais envolvidos no projeto, que o validaram como uma representação do sistema real.

O foco da aplicação do modelo computacional é na agilização do fluxo de pacientes urgentes pela implementação de uma triagem especializada com classificação de risco, antes do gargalo, de modo a priorizar os casos que demandam maior atenção naquele hospital, uma vez que os casos de emergência são prontamente atendidos, não havendo espera do paciente pelo atendimento médico.

Os resultados da simulação permitiram a análise de indicadores estratégicos como a taxa de ocupação dos médicos, quantidade de pacientes atendidos, ociosidade e sobrecarga nas clínicas e o tempo de espera por atendimento nas filas fornecendo informações fundamentais para a tomada de decisão pelos gestores.

A implementação deste passou implicou na alteração de procedimentos e rotinas na unidade de emergência, como por exemplo a disponibilização de enfermeiras profissionais de outros setores para atuar na triagem, no sistema de admissão da unidade de emergência. Do mesmo modo a alocação planejada para os médicos previu a disponibilidade destes em função da demanda por cada clínica, sendo modificada dinamicamente nos horários de pico.

## 3. Subordinar/Sincronizar todos os recursos restantes às decisões tomadas acima

A implementação deste passou implicou na alteração de procedimentos e rotinas na unidade de emergência, como por exemplo a disponibilização de enfermeiras profissionais de outros setores para atuar na triagem, no sistema de admissão da unidade de emergência. Do mesmo modo a alocação planejada para os médicos previu a disponibilidade destes em função da demanda por cada clínica, sendo modificada dinamicamente nos horários de pico.

Para evitar que a restrição seja quebrada no passo 4, e no sentido de aproveitar ao máximo os recursos no gargalo, um Aplicativo de Monitoramento da Demanda foi desenvolvido e implementado (SABBADINI, 2005). Esta ferramenta permite identificar visualmente os dias da semana e horários dentro de um dia com maior pico de procura por atendimento, assim como analisar o comportamento, hora a hora, dia a dia e semana a semana, dentro de cada mês ao longo do ano, melhorando a previsibilidade acerca da demanda e desse modo permitindo aos gestores atuar no sistema de maneira eficaz.

#### 4. Elevar a capacidade da restrição do sistema

A programação do fluxo de atendimento foi executada melhorando a alocação dos médicos nas clínicas em função da demanda e possibilitou elevar a capacidade na restrição sem a necessidade de inicialmente efetuar investimentos na aquisição de equipamentos, ampliação de instalações ou contratação de mais médicos.

5. Se a restrição se deslocar, retornar ao passo 1 e não permitir que a inércia se instale.

Dada a dinâmica dos sistemas, as soluções tende a se deteriorar ao longo do tempo. Nesse sentido a restrição pode mudar de lugar, deslocando-se para outro processo. A utilização do simulador e o Aplicativo de Monitoramento de Demanda possibilitam aos gestores do hospital o controle do dia-a-dia da unidade de emergência, contribuindo para o gerenciamento da restrição no sentido de não permitir que a inércia se instale e se transforme numa restrição do sistema.

Os resultados obtidos através do simulador e comparados com o processo real apresentaram uma diminuição no tempo médio de espera dos pacientes urgentes de 44,28 minutos para 28,86 minutos, com uma redução de tempo de 34,95%.

## 5. CONCLUSÃO

A análise do resultado dos três estudos de caso evidencia as possibilidades do uso da simulação, teoria das restrições e gestão da capacidade de forma integrada. Em todos os três casos estudados as técnicas apresentadas permitiram o gerenciamento estratégico de gargalos e a ação em termos de gestão da capacidade, que resultaram na melhoria no gerenciamento dos processos de tratamento a pacientes.

Deu-se especial ênfase a ao processo de gerenciamento da TOC, dado que tanto a simulação com os fundamentos e aplicações de gestão da capacidade são bastante consolidados. No primeiro caso, Hospital Municipal de Nova Iguaçu, em particular, partiu-se da abordagem de simulação para posteriormente através do processo de focalização identificar as ações de melhoria requeridas na gestão estratégica da capacidade. Nos demais casos, INCA e HMEHSG, partiu-se do processo de focalização da Teoria das Restrições para identificar o gargalo e posteriormente realizaram-se estudos de simulação para orientar a gestão da capacidade.

A metodologia TOC&DES apresentada demonstrou a integração das técnicas, assim como evidenciou a necessidade de em ambas as macrofases ser necessário conhecimento específico de processos e operações. O processo de focalização da Teoria das Restrições é bastante intuitivo. No mesmo sentido a macrofase de

simulação requer também especialização e conhecimento das técnicas de desenvolvimento de modelos de simulação e domínio dos fundamentos da estatística. Verificou-se também que o uso integrado e combinado das duas áreas de conhecimento reduz significativamente o esforço dispendido na identificação de gargalos em processos de serviços, permitindo o uso foca da simulação de eventos discretos na identificação das oportunidades de melhoria e análise de diferentes cenários.

O gerenciamento baseado na teoria das restrições assume uma dimensão estratégica num ambiente complexo em que se encontra o desafio de gerenciar permanentemente recursos. Ao redirecionar as ações de melhoria nos recursos gargalo, pode-se ajustar a capacidade à demanda e atuar nos pontos do sistema em que os benefícios da alteração vão repercutir no aumento da capacidade global. Em todos os casos os ganhos obtidos possuem implicações importantes como a satisfação do paciente no processo de tratamento e sua percepção em relação à qualidade dos serviços prestados. Outro aspecto a ser considerado é a melhoria da assistência prestada a estes que passam a ter seu tratamento iniciado mais rapidamente, contribuindo para a elevar a sua qualidade de vida, chances de cura e restabelecimento.

O presente método foi desenvolvido, aplicado e analisado em aplicações de serviços de saúde. Como recomendações para trabalhos futuros, sugere-se a utilização do método em aplicações em outros segmentos de serviços, assim como em manufatura industrial.

## REFERÊNCIAS

**BANKS J.; CARSON, J. S.** Discrete-event system simulation. Prentice-Hall international series in industrial and system engineering. New Jersey: Prentice Hall, 1984.

**BRAILFORD, S. C., HARPER, P. R.; PATEL, B.; PITTI M..** 2009. "An Analysis of the Academic literature on Simulation and Modelling in Health Care." J of Simulation 3(3):130-140

**BURTON, T,** GOLDRATT INSTITUTE, 2001, "TOC Case Study: Healthcare – What if my organization's goal is not to make money", disponível em <http://www.goldratt.com/for-cause/partnerperspsep2001toct.htm>

**CHAGAS, C. D.** Simulação em emergência médico-hospitalar: um estudo de caso no Hospital Municipal Miguel Couto. Dissertação (Mestrado em Engenharia de Produção) - COPPE/UFRJ, Rio de Janeiro, 2003.

**DE OLIVEIRA, MJF;TOSCANO, LNP,** Emergency information support system for Brazilian public hospital, In: M.S. Rauner and K. Heidenberger (Eds.), Quantitative approaches in health care management, Peter Lang, (2003) 235-251.

**DE OLIVEIRA, M. J. F. e ACHÃO FILHO, N.** A simulação como método de avaliação da qualidade de atendimento na emergência de um hospital municipal. In: Simpósio de Pesquisa Operacional da Marinha, 6., Simpósio de Logística da Marinha, 7., Anais. Rio de Janeiro, 2003.

**GOLDRATT, E. M.** What is this thing called Theory of Constraints and how should it be implemented? Massachusetts: North River Press, 1990.

**GOLDRATT, E.M.,**1994, **A Meta** , 17ª edição, ed. Educator, São Paulo.

**GONÇALVES, A. A.** Gestão da capacidade de atendimento em hospitais de câncer. Dissertação (Doutorado em engenharia de produção). Rio de Janeiro : COPPE/UFRJ, 2004.

**GONÇALVES A.A., DE OLIVEIRA M.J.F.,** 2004, The use of simulation to improve cancer treatment access, Accessibility and Quality of Health Services, Peter Lang, Frankfurt Main/Berlin/Bern/New York/Paris/Viena, Germany, pp.149-160.

**KERSHAW R.,** 2000, Using TOC to cure Healthcare problems”, Management Accounting Quaterly, Spring.

**MAGALHÃES, M. S.** (2006). Simulação do Sistema de Admissão de Emergência do Hospital Universitário Antônio Pedro. Dissertação de M.Sc, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

**PHIPPS B.,** 1999, “Hitting the Bottleneck’ , Health Management, United Kingdon, disponível em Goldratt Institute [www.goldratt.com/hlthmgmtmn.htm](http://www.goldratt.com/hlthmgmtmn.htm).

**ROSSETTI, M. D. et al.** Emergency department simulation and determination of optimal attending physician staffing schedules. In 1999 Winter Simulation Conference. Proceeding , Arizona 1999.

**RÉBULA, U. et. al.** Integrando técnicas e procedimentos de gestão de operações: uma aplicação em um banco comercial brasileiro de grande porte. Revista Produção. 2009.

**SABBADINI, F.S.** Gerenciamento de restrições em hospital de emergência: um estudo de caso no Hospital Municipal Henrique Sérgio Gregori. Dissertação (Mestrado em Administração e Desenvolvimento Empresarial ). Rio de Janeiro. UNESA, 2005.

**SABBADINI, F. S. ; De OLIVEIRA, M. J. F. ; GONÇALVES, A A .** Administration of constraints in a Brazilian public emergency hospital service. In: Xiolan Xie; Françoise Lorca; Éric Marcon.. (Org.). Operations Research for Health Care Delivery Engineering.. Saint-Étienne: Saint- Étienne: Publications de l'Université de Saint-Étienne, 2008, v. 33, p. 367-378.

**SLACK, N.** Administração da produção. 2. ed. São Paulo : Atlas, 2002.

**SLACK, N., CHAMBERS, R; JOHNSTON, R.** Administração da produção. 3 ed. São Paulo: Atlas, 2009.