

**SISTEMAS DE AGENDAMENTO EM SERVIÇOS DE SAÚDE AMBULATORIAIS:
REVISÃO SISTEMÁTICA DA LITERATURA**

**OUTPATIENT APPOINTMENT SCHEDULING SYSTEMS:
SYSTEMATIC LITERATURE REVIEW**

**SISTEMAS DE HORARIO EN LOS SERVICIOS DE SALUD PARA PACIENTES
EXTERNOS: REVISIÓN SISTEMÁTICA DE LA LITERATURA**

Como citar:

Oleskovicz, Marcelo & Pedroso, Marcelo Caldeira. (2023). Sistemas de agendamento em serviços de saúde ambulatoriais: revisão sistemática da literatura. Revista Gestão & Tecnologia. v. 23, n^o 2. p. 319-342, 2023.

Marcelo Oleskovicz

Doutorado em Administração pela Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP)

<https://orcid.org/0000-0002-6994-0549>

Marcelo Caldeira Pedroso

Professor Livre-Docente do Departamento de Administração da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP)

<http://orcid.org/0000-0002-2623-293X>

Editor Científico: José Edson Lara
Organização Comitê Científico
Double Blind Review pelo SEER/OJS
Recebido em 22/05/2021
Aprovado em 10/04/2023



This work is licensed under a Creative Commons Attribution – Non-Commercial 3.0 Brazil



Journal of Management & Technology, Vol. 23, n. 2, p. 319-342, 2023

Resumo

Objetivo: este estudo buscou revisar a teoria acerca de sistemas de agendamento em serviços de saúde ambulatoriais.

Metodologia/abordagem: foi realizada uma revisão sistemática de literatura a partir da base *Web of Science*, contemplando artigos em inglês e português publicados a partir de 2003.

Originalidade/relevância: ao mesmo tempo em que atualiza e organiza dados sobre a literatura relativa ao tema, o estudo aponta oportunidades relevantes para novas pesquisas.

Principais resultados: foram selecionados 65 estudos para análise, sendo observado um crescimento expressivo de pesquisas relacionadas ao tema nos últimos anos, assim como o predomínio de estudos analíticos de sistemas de agendamento com um servidor e um único estágio. Constatou-se também, dentro da segmentação adotada, diversos contextos não cobertos pelos sistemas de agendamento presentes na literatura.

Contribuições teóricas/metodológicas: entre as lacunas identificadas, destaca-se o potencial de exploração da heterogeneidade dos sistemas, especialmente a heterogeneidade associada ao período da sessão de atendimento e o recálculo de horários em função da heterogeneidade dos pacientes, visando maior liberdade de escolha de horários.

Palavras-Chave: Serviços de saúde, sistema de saúde, agendamento de consultas ambulatoriais, revisão sistemática da literatura.

Abstract

Objective: this study aimed to review the theory about appointment scheduling in outpatient health services.

Methodology/approach: a systematic literature review was carried out based on the Web of Science database, covering articles in English and Portuguese published since 2003.

Originality/relevance: at the same time that it updates and organizes data on the literature related to the topic, the study points out relevant opportunities for new research.

Main results: 65 studies were selected for analysis. There has been a significant increase in research related to the topic in recent years, as well as the predominance of analytical studies of scheduling systems with one server and a single stage. It was also found, within the classification adopted, several contexts not covered by the scheduling systems present in the literature.

Theoretical/methodological contributions: among the identified gaps, the potential for exploring the heterogeneity of the systems stands out, especially the heterogeneity associated with the consultation period and the recalculation of schedules according to the heterogeneity of patients, aiming at greater freedom in choosing schedules.

Keywords: Health care services, health care system, outpatient appointment scheduling, systematic literature review.

Resumén

Objetivo: este estudio tuvo como objetivo revisar la teoría sobre la programación de citas en los servicios de salud para pacientes ambulatorios.

Metodología/enfoque: se llevó a cabo una revisión bibliográfica sistemática basada en la base de datos Web of Science, que abarca artículos en inglés y portugués publicados desde 2003.

Originalidad/relevancia: al mismo tiempo que actualiza y organiza datos sobre la literatura relacionada con el tema, el estudio señala oportunidades relevantes para nuevas investigaciones.

Principales resultados: se seleccionaron 65 estudios para su análisis. Ha habido un aumento significativo de la investigación relacionada con el tema en los últimos años, así como el predominio de estudios analíticos de sistemas de programación con un servidor y una sola etapa. También se encontró, dentro de la clasificación adoptada, varios contextos no cubiertos por los sistemas de programación presentes en la literatura.

Contribuciones teóricas/metodológicas: entre las brechas identificadas, destaca la posibilidad de explorar la heterogeneidad de los sistemas, especialmente la heterogeneidad asociada al período de consulta y el recálculo de horarios según la heterogeneidad de los pacientes, buscando una mayor libertad en la elección de horarios.

Palabras clave: Servicios de salud, sistema de salud, programación de citas para pacientes ambulatorios, revisión sistemática de la literatura.

1 INTRODUÇÃO

Prestadores de serviços de saúde enfrentam diversos desafios para aumentar a eficiência e eficácia de seus sistemas, entre eles a crescente pressão de custos e demanda por qualidade (Safdar, Emrouznejad & Dey, 2015). No caso de serviços ambulatoriais, o bom desempenho de um sistema de agendamento é um ponto crítico na busca por equilíbrio entre utilização de recursos e qualidade de serviços (Deceuninck, Fiems & Vuyst, 2018; Lee, Heim, Sriskandarajah & Zhu, 2018).

Reduzir o tempo de espera de pacientes é um dos problemas relevantes de um prestador de serviços ambulatoriais (Gupta & Denton, 2008). Em geral, demanda-se que cada paciente obtenha antecipadamente um horário de agendamento, compareça ao prestador no horário agendado e então receba o serviço programado. A natureza estocástica de variáveis como tempos de serviço, pontualidade dos pacientes, *no-shows* e cancelamentos produzem uma grande variedade de situações complexas e muito esforço tem sido direcionado a sistemas de agendamento em termos de pesquisa (Morikawa & Takahashi, 2017; Xiao, Dong, Li & Sun, 2017).

Os problemas de agendamento ocorrem em vários contextos operacionais, como transporte, manufatura e serviços (Vanden Bosch, Dietz & Simeoni, 1999) e muitas décadas de

pesquisa têm sido dedicadas a este tema (Leung, 2004). Entretanto, devido às suas características únicas, estas técnicas não podem ser simplesmente replicadas para agendamento de serviços de saúde. Mesmo dentro deste setor, cenários diversos exigem soluções específicas. Choi e Banerjee (2016), por exemplo, citam as diferenças entre serviços ambulatoriais e cirurgias que impactam sistemas de agendamento. Enquanto nos primeiros o número de pacientes atendidos é maior, em cirurgias os procedimentos são mais longos e sujeitos a maior variação de tempo.

As atuais pesquisas sobre agendamentos ambulatoriais fornecem muitas técnicas úteis, porém o tema ainda precisa de aprofundamento para o tratamento de problemas ainda mais realistas (Lee, Heim, Sriskandarajah & Zhu, 2018). Muitos estudos não permitem generalização ou contemplam um conjunto muito restrito de premissas (Cayirli & Veral, 2003; Kuiper, Kemper & Mandjes, 2015).

Os primeiros estudos sobre sistemas de agendamento de serviços ambulatoriais tiveram início na década de 1950, com os trabalhos de Bailey (1952) e Lindley (1952). Desde então, o número de pesquisas sobre este assunto tem sido crescente. Uma abrangente revisão da teoria foi elaborada no trabalho seminal de Cayirli e Veral (2003), a qual tem sido largamente empregada como referência em pesquisas. Além de condensar a teoria, este estudo traz uma interessante estrutura, a qual classifica os aspectos referentes a sistemas de agendamento, bem como contribui para padronização da sua terminologia.

Outra referência é o trabalho de Gupta e Denton (2008), o qual aponta desafios e oportunidades em sistemas de agendamento no setor de saúde, e não somente em serviços ambulatoriais, contemplando também o estado da arte acerca do tema. Mais recentemente, Ahmadi-Javid et al. (2017) realizaram uma detalhada revisão de literatura sobre sistemas de agendamento ambulatorial, restringindo a pesquisa a estudos com uso de otimização publicados a partir de 2003, dada a relevância da revisão de Cayirli e Veral (2003).

Nesse contexto, este estudo contempla uma revisão sistemática da literatura acerca de sistemas de agendamento ambulatorial. Mais precisamente, busca responder à seguinte pergunta de pesquisa: Quais são os sistemas de agendamento para serviços ambulatoriais existentes na literatura e como ela os classifica?

Para tanto, considerando o trabalho de Cayirli e Veral (2003) como referência das publicações até 2003, este artigo abrange pesquisas publicadas a partir daquele ano e se baseia em sua estrutura para a classificação dos estudos selecionados.

2 MÉTODO, RESULTADOS E DISCUSSÃO DO ESTUDO

Este estudo consiste em uma revisão sistemática de literatura acerca de sistemas de agendamento de serviços de saúde ambulatoriais. A identificação dos estudos a serem revisados se deu por meio da base *Web of Science*, sendo considerados na busca os termos (“appointment” OR “outpatient appointment”) AND (“scheduling”) AND (“health” OR “health care” OR “healthcare”) entre os tópicos dos registros. Como um primeiro filtro, foram mantidos somente estudos em inglês e português. Na sequência, foram selecionados estudos na forma de artigo. Adicionalmente, dada a abrangência da revisão de literatura seminal de Cayirli e Veral (2003), a seleção foi restrita a artigos publicados a partir de 2003.

Concluída a triagem, foi realizada a fase de elegibilidade do material, iniciando-se com a leitura de títulos e resumos e consequente exclusão de itens não compatíveis com os objetivos da pesquisa. Ainda como parte desta fase, foram analisados os textos dos artigos remanescentes, com nova exclusão de estudos não relacionados. Dentro da elegibilidade buscou-se manter somente artigos contemplando pelo menos um modelo de sistema de agendamento, com abordagem quantitativa. Esta fase também desconsiderou sistemas com agenda aberta, onde não há definição de horários de agendamento. Este critério tem base no fato de esta abordagem contemplar elementos de decisão específicos, embora as medidas de desempenho sejam usualmente as mesmas de sistemas com horários agendados.

A Figura 1 demonstra o fluxo aplicado para seleção dos estudos. Como resultado deste processo, 65 artigos foram selecionados para análise qualitativa e quantitativa. Esta análise é apresentada nas seções seguintes, as quais se baseiam na estrutura proposta por Cayirli e Veral (2003).

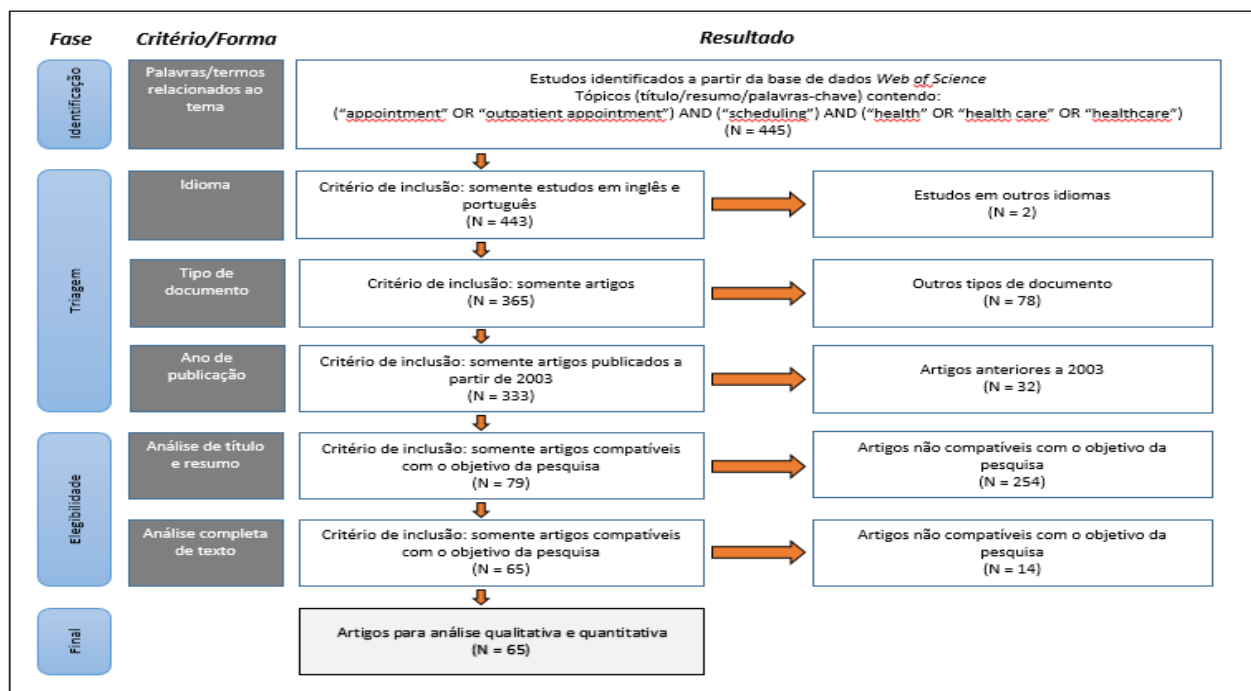


Figura 1 – Fluxo do processo de Revisão Sistemática da Literatura (RSL)

Fonte: Os autores.

A Tabela 1 traz a segmentação destes 65 estudos por período de publicação, permitindo constatar o recente crescimento no número de pesquisas sobre o tema.

Tabela 1

Número de Estudos por Período

Período	Número de Estudos
2003 a 2006	3
2007 a 2010	4
2011 a 2014	26
2015 a Mar-2018	32
Total	65

Fonte: Os autores.

Os resultados da revisão sistemática da literatura são apresentados a seguir, agrupados nos seguintes tópicos: fatores ambientais; indicadores de desempenho; desenvolvimento de um sistema de agendamento; e métodos de pesquisa em sistemas de agendamento ambulatoriais. Estes são discutidos a seguir.

Fatores Ambientais

Serviços ambulatoriais podem ser considerados sistemas de filas, onde os parâmetros devem ser configurados conforme as especificidades de cada cenário. O caso mais simples é o de um sistema onde não há atrasos dos pacientes, com um único profissional para atendimento e tempos de serviço estocásticos (Cayirli & Veral, 2003). Tanto os serviços quanto os pacientes atendidos por um prestador apresentam características que impactam a configuração de um sistema de agendamento, comumente denominadas fatores ambientais. Os seguintes fatores ambientais são tradicionalmente considerados no agendamento de serviços ambulatoriais: (1) tipos de agendamento; (2) número de estágios; (3) número de servidores; (4) número de atendimentos por sessão; (5) processo de chegadas; (6) tempo de serviço; e (6) atrasos e interrupções do servidor. Estes são apresentados a seguir.

Tipos de Agendamento

Dois tipos de agendamento podem ser encontrados na literatura: sequencial e simultâneo. No primeiro os pacientes são agendados no momento da solicitação, enquanto no tipo simultâneo os agendamentos são realizados somente depois que todas as solicitações foram feitas (Zacharias & Pinedo, 2014). Pérez et al (2013) apresentam o tipo sequencial estocástico, como uma extensão do agendamento sequencial, o qual considera o potencial de solicitações a serem recebidas após cada solicitação de agendamento.

Wang et al. (2018) apresentam um modelo que, conforme o estado do sistema no momento do agendamento, oferece ao paciente um determinado conjunto de horários, de modo a maximizar a receita do prestador.

Embora a prática do agendamento sequencial seja mais comum, o tipo simultâneo predomina na literatura, uma vez que sua modelagem é menos complexa (Ahmadi-Javid, Jalali & Klassen, 2017). Em alguns estudos, como o de Zacharias e Pinedo (2014), os resultados obtidos a partir de um caso simultâneo são utilizados para analisar o caso correspondente do tipo sequencial.

Na medida em que os sistemas de agendamento eletrônico têm evoluído rapidamente (Wang e Gupta, 2011), a importância do agendamento simultâneo tem crescido, uma vez que as solicitações são coletadas via *e-mail* ou portal na *web* e, na sequência, os pacientes são informados sobre os horários de agendamento (Ahmadi-Javid, Jalali & Klassen, 2017).

Existem estudos recentes contemplando sistemas sequenciais, entretanto existe espaço para aprofundamento, em especial a abordagem de casos mais realistas.

A Tabela 2 (apêndice) contempla os tipos de agendamento dos estudos analisados.

Número de Estágios

A maior parte dos estudos modela um sistema de um estágio, onde os pacientes agendam um único serviço. Nos poucos estudos com cenários multiestágio (Hahn-Goldberg et al. 2014; Tsai & Teng, 2014; Lin, 2015; Kuiper & Mandjes, 2015), identificados no Quadro 1 (apêndice), os horários de cada estágio são agendados separadamente. Existe, portanto, oportunidade para futuras pesquisas com cenários de múltiplos estágios.

Número de Servidores

O número de servidores para atendimento da demanda é um fator importante no desenvolvimento de um sistema de agendamento, impactando custos e qualidade dos serviços (Ahmadi-Javid, Jalali & Klassen, 2017). Na literatura revisada há predomínio de estudos com um único servidor. Segundo Cayirli e Veral (2003), a motivação reside no fato de os pacientes terem preferência quanto a atendimento por um determinado médico, para o caso de consultas particulares.

Os principais recursos identificados na literatura revisada incluem equipes, salas, leitos/cadeiras e equipamento médico. Ahmadi-Javid et al. (2017) destacam os seguintes ambientes para os casos em que os servidores consistem em equipamentos: diferentes conjuntos de equipamentos idênticos, conjunto único de equipamentos idênticos e diferentes conjuntos com um único equipamento.

Embora o número de servidores possa impactar o desempenho de serviços ambulatoriais, conforme já mencionado, este parâmetro é um dado de entrada na grande parte

das pesquisas analisadas. Torná-lo um fator de decisão em modelos futuros, portanto, configura-se como uma oportunidade. O número de servidores das pesquisas revisadas é demonstrado no Quadro 1 (apêndice).

Número de Atendimentos por Sessão

Desde os primeiros estudos sobre sistemas de agendamento, como o de Welch e Bailey (1952), o número de atendimentos em uma sessão é considerado fator relevante no que diz respeito a desempenho. Pesquisas que buscam otimizar o número de atendimentos por sessão normalmente têm por objetivo a minimização dos tempos de espera de pacientes e horas-extras do servidor. Nos modelos de Cayirli et al. (2012) e Cayirli e Yang (2014), os intervalos entre agendamentos são definidos em função desse parâmetro.

O número de atendimentos em uma sessão impacta, ainda, o estado da fila de um sistema de agendamento e tem uma grande variação entre os trabalhos analisados. A pesquisa de Kuiper et al. (2015), por exemplo, estuda tanto estados estáveis quanto transientes. Para os cenários avaliados, o comportamento transiente é notado em sessões com até 25 atendimentos.

Fatores como absentéismo (*no-show*) e presença de pacientes não-agendados (*walk-ins*) também influenciam o número de atendimentos, na medida em que este pode exercer uma forma de compensação. O Quadro 1 (apêndice) aponta o número de atendimentos por sessão dos estudos analisados.

Processo de Chegadas

Segundo Cayirli e Veral (2003), um conjunto de fatores compõe as características do processo de chegadas de pacientes, impactando o desempenho do sistema de agendamento. Um destes fatores é a pontualidade do paciente. Os autores definem a falta de pontualidade como a diferença entre o horário agendado e o horário real da sua chegada, tanto na forma de atraso quanto de antecipação. A maioria dos autores que incluem a pontualidade em seus estudos contempla apenas os atrasos, considerando chegadas antecipadas como pontuais. Klassen e Yoogalingam (2014) analisam alternativas de tamanhos de blocos de pacientes e intervalos entre horários de agendamento para diferentes cenários de pontualidade, identificando as mais

efetivas. Cayirli e Yang (2014) conduzem uma análise de sensibilidade para avaliar os efeitos da pontualidade no desempenho de seu modelo.

Outro fator relevante é o *no-show*. Em seu estudo sobre desafios e oportunidades em agendamentos, Gupta e Denton (2008), assim como Ahmadi-Javid et al. (2017), destacam este fenômeno como um ponto a ser explorado em maior profundidade. Os casos de cancelamentos com baixa antecedência (*late cancellations*), em que não há possibilidade de se agendar outro paciente no horário cancelado, são usualmente tratados como *no-shows* na literatura. O estudo de Deceuninck et al. (2018), entretanto, faz a seguinte distinção: em caso de *late cancellations* o paciente seguinte, caso tenha chegado antecipadamente, terá seu atendimento antecipado, enquanto no caso de *no-show*, deverá aguardar seu horário agendado.

Um terceiro fator a constituir o processo de chegadas é a aceitação de *walk-ins*, ou pacientes sem agendamento, classificados por Cayirli e Veral (2003) em regulares e urgentes, conforme o nível de prioridade de atendimento. Assim como *no-shows* podem ser compensados por meio de *overbooking* ou redução de intervalos entre agendamentos, o efeito de *walk-ins* pode ser mitigado mediante o bloqueio de horários na agenda ou um maior intervalo entre agendamentos. Morikawa e Takahashi (2017) tratam destas alternativas, alocando *walk-ins* em uma agenda de blocos múltiplos, previamente definida.

Considerar simultaneamente *no-shows* e *walk-ins* melhora o desempenho de um sistema de agendamento, segundo Ahmadi-Javid et al. (2017). Cayirli et al. (2012) e Cayirli e Yang (2014) sugerem um procedimento que ajusta a média e o desvio padrão dos tempos de serviço com base nas probabilidades de *no-shows* e *walk-ins*, buscando reduzir seus efeitos negativos.

A presença de *walk-ins* implica maior complexidade de modelagem, sendo esta a provável razão pela qual a maioria das pesquisas com abordagem de otimização não inclui este parâmetro em seus modelos, configurando uma lacuna na literatura (Ahmadi-Javid, Jalali & Klassen, 2017). As características referentes a pontualidade, *no-shows* e *walk-ins* dos trabalhos revisados podem ser identificados no Quadro 1 (apêndice).

Tempo de Serviço

O tempo de serviço pode ser definido como a soma de todos os tempos em que o paciente está demandando atenção de um servidor, impedindo este de atender outros pacientes (Cayirli & Veral, 2003), podendo ser tanto um valor estocástico, com uma distribuição estatística, como um valor determinístico (Gupta & Denton, 2008). A modelagem de problemas de agendamento com tempos de serviço estocásticos é amplamente explorada na literatura (Erdogan & Denton, 2013; Ahmadi-Javid, Jalali & Klassen, 2017).

Uma alta variabilidade dos tempos de serviço afeta negativamente tanto o tempo de espera dos pacientes como a ociosidade do prestador de serviços (Denton & Gupta, 2003; Cayirli et al, 2012). Distribuições de probabilidade que representam situações reais normalmente implicam complexidade para modelos analíticos. Kuiper et al. (2015) aplicam técnica de aproximação com distribuições faseadas para tempos de serviço, obtendo bons resultados, com baixo esforço computacional. Em seu modelo de otimização, Chakraborty et al. (2013) demonstram que a aplicação de distribuições gamma para tempos de serviço também proporcionam baixo tempo de processamento.

A maior parte das pesquisas assume distribuições idênticas, havendo uma tendência de estudos com distribuições heterogêneas, onde o tempo de serviço depende do tipo de paciente ou serviço (Ahmadi-Javid, Jalali & Klassen, 2017). O Quadro 1 (apêndice) demonstra como os tempos de serviço são considerados nos estudos analisados.

Atrasos e Interrupções do Servidor

O tempo de espera dos pacientes é significativamente sensível a atrasos do servidor. Caso o primeiro serviço não seja iniciado pontualmente, observa-se um acúmulo de atrasos que se estende por toda a sessão de atendimento (Cayirli & Veral, 2003). Entre os poucos estudos que tratam deste tema, a pesquisa de Klassen e Yoogalingam (2013) conclui que quanto maior o atraso, mais a sessão deve ter seu início postergado e os intervalos entre atendimentos devem ser mais curtos. Outro fator relacionado ao servidor é a interrupção no atendimento. No caso de médicos, por exemplo, inclui todas as atividades durante uma sessão que podem requerer a atenção do profissional, como interações com a equipe e telefonemas (Cayirli & Veral, 2003).

Para o caso de interrupções, Klassen e Yoogalingam (2013) recomendam intervalos mais longos entre serviços no meio da sessão. Luo et al (2012) simulam distribuições de Poisson para a frequência de interrupções do servidor e afirmam que estas causam maior impacto quando não são uniformes ao longo de uma sessão de atendimento. As pesquisas que consideram interrupções do servidor estão apontadas no Quadro 1 (apêndice).

Indicadores de Desempenho

Existe uma série de critérios utilizada na literatura para avaliar sistemas de agendamento, os quais buscam equilibrar objetivos conflitantes das partes interessadas, frequentemente servidores e pacientes (Ahmadi-Javid, Jalali & Klassen, 2017). Os resultados nas pesquisas são medidos usualmente em termos de tempo de espera de pacientes, tempo de ociosidade do servidor, horas-extras do servidor, número de pacientes atendidos e número de pacientes rejeitados pelo sistema (Cayirli & Veral, 2003; Ahmadi-Javid et al. 2017). A alternativa de maior frequência na literatura (p. ex.: Lee et al., 2018; Jiang et al., 2017, Xiao et al., 2017; Salzarulo et al., 2016) consiste em atribuir pesos aos custos relacionados a espera do paciente (C_p), ociosidade (C_d) e horas-extras do servidor (C_o). Deste modo, o objetivo seria atingir um custo total esperado mínimo, representado na seguinte equação:

$$\text{Mín } E(TC) = E(W)C_p + E(I)C_d + E(O)C_o$$

Em geral assume-se que $C_d > C_p$, uma vez que o C_d abrange não somente o custo do médico, mas também das instalações (Cayirli et al., 2012). Alguns estudos não consideram o tempo de espera do paciente como um critério de desempenho. Tendo como premissa custos diferentes para médicos de uma clínica, o modelo de Wang et al. (2018) tem como objetivo maximizar a receita total do prestador. Kim e Giachetti (2016) consideram o lucro do servidor, baseado em receita, custos fixos e penalidades para hora-extra. Turkcan et al. (2011) consideram tempos de espera do paciente e horas-extras do servidor como restrições em vez de parte da função objetivo de seu modelo.

Morikawa et al. (2018) aplicam o conceito de *clearing function*, amplamente utilizado em problemas referentes a cadeia de suprimentos, estabelecendo relação entre tempos de espera

e utilização do servidor. Qi (2016) propõe um indicador de insatisfação relacionado ao atraso no atendimento, o qual considera a frequência e intensidade de atrasos acima de um determinado limite, e busca balancear o nível de serviço entre pacientes. Riise et al. (2016) buscam equilibrar o número de pacientes agendados e o tempo de espera indireto, calculado pela diferença entre o dia da solicitação e o dia do atendimento. Empregando a técnica de *overbooking*, o modelo de Creps e Lofti (2017) considera o número de conflitos, representado pelo número de pacientes que comparecem em um mesmo horário, e a receita do servidor. Gocgun e Puterman (2014), Schuetz e Kolisch (2013) e Kim e Giachetti (2016) são exemplos de pesquisas que consideram penalidade para a rejeição de pacientes entre os critérios de desempenho. O Quadro 1 (apêndice) resume os indicadores de desempenho dos estudos revisados.

Desenvolvimento de um Sistema de Agendamento

Cayirli e Veral (2003) consideram um sistema de agendamento como uma série de decisões, divididas em três conjuntos: regras de agendamento, uso de classificação de pacientes (quando aplicável) e ajustes para mitigar os efeitos negativos de *no-shows* e *walk-ins*.

Regras de Agendamento

Sessões de atendimento são usualmente divididas em blocos, sendo que cada bloco equivale a um horário agendado. As regras para agendamento envolvem o número de pacientes em cada bloco e o intervalo entre blocos. Uma vez que tanto o tamanho dos blocos e os intervalos podem ser fixos ou variáveis em uma mesma sessão, várias combinações podem ser feitas com estes parâmetros (Cayirli & Veral, 2003).


A forma mais elementar de um sistema de agendamento é o bloco simples, o qual consiste em agendar todos os pacientes para que cheguem como um bloco no início da sessão, sendo atendidos por ordem de chegada. Obviamente esta regra implica tempos de espera excessivos para pacientes, ao mesmo tempo em que proporciona uma alta ocupação do servidor.

Esta era a prática comum na década de 1950, a partir da qual tiveram início as pesquisas em

agendamento de serviços ambulatoriais. Deste modo, observa-se que estes primeiros estudos exploram as vantagens do agendamento individual, sendo os pioneiros na transição de sistemas de bloco simples para sistemas de bloco individual (Lindley, 1952; Bailey, 1952; Fry, 1964; Johnson & Rosenfeld, 1968; Rockart & Hofmann, 1969). Bailey (1952) propôs em seu estudo um bloco inicial com dois pacientes e intervalos fixos entre agendamentos, apontando os impactos para pacientes e servidores. Esta regra ficou conhecida como Regra de Bailey, a qual ainda é amplamente considerada na literatura.

O modelo de otimização e as heurísticas propostas por Choi e Banerjee (2016) utilizam intervalos fixos com blocos variáveis. Os cenários analisados contemplam variações sobre a regra de Bailey. Neste estudo estas variações são denominadas regras de Bailey periódicas, uma vez que não somente o bloco inicial, mas também blocos intermediários (distanciados por determinados períodos) contêm mais pacientes que os demais blocos. Conforme Zacharias e Pinedo (2014), blocos com mais de um paciente devem ser agendados para o início da sessão na maioria das regras de otimização analisadas. Em seu estudo analítico sobre intervalos variáveis, Wang (1993) concluiu que intervalos ótimos são crescentes no início de uma sessão de atendimento, decrescendo a partir da sua metade em direção ao final. Devido a este comportamento, este perfil de intervalos é referenciado como “formato de domo” e é aplicado em heurísticas como as de Denton e Gupta (2003), Cayirli et al. (2012) e Cayirli e Yang (2014) e estudos com otimização, como o de Kuiper e Mandjes (2015).

Luo et al. (2012) concluem que, para altas taxas de interrupção do servidor, intervalos crescentes ao longo da sessão trazem melhores resultados quando comparados com o “formato de domo”. Além da interrupção no servidor, Klassen e Yoogalingam (2013), também focam o seu atraso, concluindo que intervalos maiores entre agendamentos podem mitigar os efeitos das interrupções, enquanto os atrasos do servidor são compensados mediante deslocamento dos horários de atendimento em direção ao final da sessão. Su e Shih (2003) simulam diversos valores de intervalos fixos e várias formas de alocar horários para pacientes agendados e *walk-ins*, medindo separadamente o impacto nestes dois tipos de pacientes. No estudo de Chakraborty et al. (2013), os horários não são pré-definidos. O sistema escolhe um horário conforme critérios de desempenho e os períodos disponíveis são redefinidos dinamicamente, à medida que os

 **Journal of Management & Technology, Vol. 23, n. 2, p. 319-342, 2023**

pacientes são agendados. A pesquisa conclui que intervalos ótimos dependem da variabilidade dos tempos de serviço, bem como o comportamento dos *no-shows*. A pontualidade do paciente também exerce influência sobre os intervalos ótimos entre serviços (Klassen & Yoogalingam, 2014; Tai & Williams. 2012).

Na Tabela 2 (apêndice) constam os tipos de intervalo e blocos de cada trabalho revisado.

Heterogeneidade e Classificação de Pacientes

Em uma grande parte dos estudos os pacientes são considerados homogêneos e agendados por ordem de solicitação. Entretanto, quando fatores como tempo de serviço e padrões de chegada permitem agrupá-los em classes, levanta-se a possibilidade de aprimorar o sistema de agendamento com o uso destas características (Cayirli & Veral, 2003). Zacharias e Pinedo (2014) afirmam que a heterogeneidade de pacientes tem um impacto significativo no desempenho de um sistema de agendamento.

Segundo Cayirli e Veral (2003), a classificação de pacientes em agendamentos ambulatoriais é usualmente explorada com dois objetivos: sequenciar pacientes no momento do agendamento e/ou ajustar os intervalos entre agendamentos com base nas diferentes características das diferentes classes de pacientes. Como a agenda comumente precisa ser feita de modo antecipado e as solicitações de agendamentos são dinâmicas, o uso da classificação de pacientes é um tanto limitada. Uma aplicação realista da classificação demanda que esta seja feita com um número de grupos gerenciável e que estes sejam alocados em janelas de tempo pré-determinadas ao solicitar um agendamento (Cayirli & Veral, 2003).

Em sistemas com pacientes homogêneos o sequenciamento é indiferente, portanto o foco são os horários de atendimento e os modelos têm como objetivo definir horários onde os pacientes são encaixados à medida que solicitam um serviço. Em sistemas com pacientes heterogêneos, entretanto, muitas pesquisas lidam com o sequenciamento, buscando otimizá-lo. Dada a complexidade inerente a este tipo de problema, a maioria dos estudos aplica alguma regra para esta definição (Ahmadi-Javid, Jalali & Klassen, 2017).

O critério mais comumente utilizado para obtenção de sequências ótimas é a ordem crescente de variância dos tempos de serviço dos pacientes (MAK; RONG; ZHANG, 2014),

sendo que sua eficácia depende do número de pacientes no sistema e a distribuição dos seus tempos de serviço (Ahmadi-Javid, Jalali & Klassen, 2017). No estudo de Berg et al. (2014), além deste critério, o sequenciamento de pacientes em ordem crescente de taxa de *no-shows* também contribui para melhorar o desempenho do seu modelo. Bhattacharjee e Ray (2016) incluem ainda a pontualidade como variável em sua pesquisa.

Para os casos de agendamento simultâneo, onde a agenda é definida após recebidas todas as solicitações de agendamento, predominam modelos de otimização em que tanto o sequenciamento quanto os horários são definidos em função de características dos pacientes. Kong et al. (2013), Gocgun e Puterman (2014) e Huang e Bach (2016) distinguem pacientes novos de pacientes de retorno, aplicando este critério para sequenciamento e definição de horários de atendimento. Salzarulo et al. (2016) utilizam o tempo de serviço e seu erro padrão para definir os agendamentos.

Em sistemas de agendamento sequencial, em que os pacientes têm seus horários de atendimento definidos à medida que as solicitações são feitas, normalmente são aplicadas heurísticas para estabelecer horários para pacientes homogêneos, sendo que estes podem escolher qualquer horário livre. Já para o caso de heterogeneidade dos pacientes, esta é utilizada para definir este horário ou previamente reservar faixas de horários para diferentes classes. Schuetz e Kolisch (2013) distinguem casos ambulatoriais de internações, entre outros parâmetros, para reservar janelas de atendimento para as diferentes classes de pacientes. Cayirli e Yang (2014) determinam horários específicos de atendimento para classificação de pacientes baseada em média e desvio-padrão dos tempos de atendimento, probabilidade de *no-shows* e probabilidade de *walk-ins*.

Para o caso de abordagens de otimização, sequências ótimas exatas ainda são desconhecidas para casos com três pacientes ou mais, devido à complexidade matemática em se determinar simultaneamente sequências e horários (Ahmadi-Javid, Jalali & Klassen, 2017).

Observa-se que na literatura não há estudos que exploram a heterogeneidade de pacientes associadas a períodos da sessão. Oleskovicz et al. (2014), por exemplo, apontam diferentes taxas de *no-show* entre os períodos da manhã e da tarde em sessões de atendimento de diversas especialidades. Isto possibilitaria utilizar a heterogeneidade para modelos

sequenciais com horários pré-definidos, sem necessidade de informações antecipadas dos pacientes, reduzindo as restrições de escolha de horários.

Nas pesquisas com agendamento sequencial com horários pré-definidos, estes não são recalculados à medida que os pacientes são agendados e suas características são conhecidas. O trabalho de Hahn-Goldberg et al. (2014) é uma exceção, entretanto seu modelo assume tempos de serviço determinísticos, o que é considerado uma premissa não realista pela literatura (Hong, Shang, Arumugam & Yusuff, 2013). No modelo sequencial de blocos múltiplos de Yan et al. (2015) os pacientes são divididos em grupos, conforme suas probabilidades de *no-show*, e a cada agendamento o recálculo se dá no número de vagas disponíveis por bloco, sem que os horários dos blocos sejam alterados.

Um resumo das características de heterogeneidade e opções do paciente dos estudos analisados pode ser verificado na Tabela 2 (apêndice).

Ajustes para *No-shows* e *Walk-ins*

Uma forma de se compensar os efeitos negativos da incerteza do processo de chegadas é o ajuste dos intervalos entre serviços. Para *no-shows* esta compensação se dá por meio de redução dos intervalos, enquanto no caso de *walk-ins* aplica-se um incremento. Como alternativa, a literatura contempla o agendamento de um número de pacientes superior à capacidade teórica (*overbooking*) para mitigar o efeito de *no-shows* ou a reserva de horários livres na agenda para acomodar *walk-ins* (Cayirli & Veral, 2003).

O modelo analítico de LaGanga e Lawrence (2012) obtém ganhos de desempenho mediante aplicação de *overbooking* em um modelo de intervalos fixos. Os autores concluem que os resultados, no entanto, não podem ser generalizados, uma vez que dependem do cenário estudado. Também empregando *overbooking*, Creps e Lofti (2017) consideram os conflitos de agenda como uma das métricas de desempenho, com base em ocorrências de um ou mais pacientes comparecendo em um mesmo horário. Morikawa e Takahashi (2017) desenvolvem um sistema em que os *walk-ins* são alocados em blocos reservados conforme suas características. Anderson et al. (2015) aplicam um fator de ajuste para os intervalos fixos de seu modelo, de modo a compensar *no-shows*. Nos trabalhos de Cayirli et al. (2012) e Cayirli e

Yang (2014) são aplicados ajustes de média e variância dos tempos de serviço, tratando-se *no-shows* como atendimentos de tempo nulo e *walk-ins* como atendimentos com o dobro do tempo médio.

Métodos de Pesquisa em Sistemas de Agendamento Ambulatoriais

Adotando a classificação de Cayirli e Veral (2003) os estudos sobre sistemas de agendamento podem ser segmentados em três métodos de pesquisa: estudos analíticos, simulações e estudos de caso.

A complexidade para o desenvolvimento de modelos analíticos para sistemas de agendamento se deve ao número de variáveis e suas distribuições de probabilidade, o que torna intratável a maioria dos problemas (Ahmadi-Javid, Jalali & Klassen, 2017). Deste modo, abordagens analíticas assumem, em sua maioria, premissas não-realistas que limitam sua aplicação, como, por exemplo, tempos de serviço com distribuições de probabilidades exponenciais (Cayirli, Yang & Quek, 2012). O estudo de Kuiper et al. (2015) ilustra os esforços em se contornar a questão das distribuições de probabilidade. O trabalho utiliza distribuições faseadas, analiticamente tratáveis, sem perda significativa de desempenho.

Sistemas de agendamento podem ser considerados sistemas de filas. Estas filas geralmente são transientes, não atingindo um estado de estabilidade, sendo este um fator que também contribui para a dificuldade de abordagens analíticas (Kuiper, Kemper & Mandjes, 2015). Por serem em muitos casos tratáveis somente em pequenas instâncias, modelos analíticos muitas vezes são utilizados no desenvolvimento de heurísticas para problemas de maior escala (Cayirli & Veral, 2003). Entretanto, a recente evolução nos métodos analíticos tem permitido a obtenção de resultados ótimos para instâncias cada vez maiores (Ahmadi-Javid, Jalali & Klassen, 2017).

Dadas estas complexidades, são desenvolvidos estudos com o uso de simulações, muitas vezes combinadas com heurísticas e aproximações (Gupta & Denton, 2008). Segundo Cayirli e Veral (2003), uma vantagem do uso de simulações em relação a abordagens analíticas é a possibilidade de se modelar sistemas de filas complexos e representar variáveis ambientais, como atributos relacionados a pacientes ou servidores. As pesquisas conduzem experimentos

de simulação para avaliar o desempenho de diferentes sistemas de agendamento e/ou compreender a relação entre diversos fatores ambientais e medidas de desempenho.

Em estudos de caso, os pesquisadores analisam um ambiente em particular, fazem recomendações para aprimoramento do sistema existente e, em alguns casos, avaliam os resultados de implantações (Cayirli & Veral, 2003). Muito embora os estudos de caso ofereçam recomendações valiosas com base em situações reais, sua maior desvantagem é a falta de generalização. O Quadro 1 (apêndice) aponta os métodos de pesquisa empregados em cada um dos estudos revisados.

3 CONSIDERAÇÕES FINAIS

Sistemas de agendamento são um importante elemento de serviços de saúde ambulatoriais, uma vez que representam o elo entre eficiência na utilização de recursos e qualidade no atendimento. Sua importância tem crescido à medida que aumenta a ênfase em medicina preventiva e tempos de internação mais curtos.

Este estudo buscou revisar a literatura recente acerca de sistemas de agendamento ambulatoriais. Foram analisados 65 artigos, sendo os resultados segmentados em: fatores ambientais, indicadores de desempenho, desenvolvimento de sistemas de agendamento e métodos de pesquisa.

Observou-se um aumento no interesse em estudos sobre o tema, mediante um crescimento no número de pesquisas nos últimos anos. Também foi possível constatar a predominância de abordagens analíticas como método de pesquisa, assim como sistemas de um estágio e um único servidor.

Quanto às oportunidades identificadas, duas não estão contempladas em revisões de literatura recentes e, de um modo geral, estão relacionadas à heterogeneidade dos pacientes e à sua liberdade para escolha de um horário na agenda. A primeira seria a ausência de estudos contemplando a heterogeneidade de pacientes associada a períodos da sessão de atendimento, a qual poderia ser utilizada em modelos sequenciais com horários pré-definidos, sem necessidade de informações antecipadas dos pacientes, reduzindo a imposição de horários ou faixas de horários para diferentes classes.

A segunda diz respeito ao recálculo de horários da agenda. Nas pesquisas com agendamento sequencial com horários pré-definidos, estes não são recalculados à medida que os pacientes são agendados e suas heterogeneidades são identificadas. O único estudo com esta abordagem adota tempos de serviço determinísticos, sendo esta uma premissa dificilmente encontrada em contextos reais.

Como limitação, este estudo não apresenta aprofundamento em alguns tópicos como, por exemplo, métodos de pesquisa e horizonte de agendas. Dada a abrangência do tema, diferentes restrições de escopo de pesquisa permitiriam novos estudos com foco em aspectos que demandam maior detalhamento. Por fim, a fonte de dados poderia ser estendida para outras bases, além da *Web of Science*.

REFERÊNCIAS

- Ahmadi-Javid, A., Jalali, Z., & Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1), 3-34.
- Akhavizadegan, F., Ansarifar, J., & Jolai, F. (2017). A novel approach to determine a tactical and operational decision for dynamic appointment scheduling at nuclear medical center. *Computers & Operations Research*, 78, 267-277.
- Alrefaei, M. H., & Diabat, A. (2015). Modelling and optimization of outpatient appointment scheduling. *RAIRO-Operations Research*, 49(3), 435-450.
- Anderson, K., Zheng, B., Yoon, S. W., & Khasawneh, M. T. (2015). An analysis of overlapping appointment scheduling model in an outpatient clinic. *Operations Research for Health Care*, 4, 5-14.
- Bailey, N. T. (1952). A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14(2), 185-199.
- Berg, B. P., Denton, B. T., Erdogan, S. A., Rohleder, T., & Huschka, T. (2014). Optimal booking and scheduling in outpatient procedure centers. *Computers & Operations Research*, 50, 24-37.
- Bhattacharjee, P., & Ray, P. K. (2016). Simulation modelling and analysis of appointment system performance for multiple classes of patients in a hospital: a case study. *Operations Research for Health Care*, 8, 71-84.
- Cayirli, T., & Gunes, E. D. (2014). Outpatient appointment scheduling in presence of seasonal walk-ins. *Journal of the Operational Research Society*, 65(4), 512-531.
- Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: a review of literature. *Production and Operations Management*, 12(4), 519-549.

- Cayirli, T., & Yang, K. K. (2014). A universal appointment rule with patient classification for service times, no-shows, and walk-ins. *Service Science*, 6(4), 274-295.
- Cayirli, T., Veral, E., & Rosen, H. (2008). Assessment of patient classification in appointment system design. *Production and Operations Management*, 17(3), 338-353.
- Cayirli, T., Yang, K. K., & Quek, S. A. (2012). A universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management*, 21(4), 682-697.
- Chakraborty, S., Muthuraman, K., & Lawley, M. (2010). Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*, 42(5), 354-366.
- Chakraborty, S., Muthuraman, K., & Lawley, M. (2013). Sequential clinical scheduling with patient no-show: The impact of pre-defined slot structures. *Socio-Economic Planning Sciences*, 47(3), 205-219.
- Chen, P. S., Robielos, R. A. C., Palaña, P. K. V. C., Valencia, P. L. L., & Chen, G. Y. H. (2015). Scheduling patients' appointments: Allocation of healthcare service using simulation optimization. *Journal of Healthcare Engineering*, 6(2), 259-280.
- Choi, S. S., & Banerjee, A. A. (2016). Comparison of a branch-and-bound heuristic, a newsvendor-based heuristic and periodic Bailey rules for outpatients appointment scheduling systems. *Journal of the Operational Research Society*, 67(4), 576-592.
- Creps, J., & Lotfi, V. (2017). A dynamic approach for outpatient scheduling. *Journal of Medical Economics*, 20(8), 786-798.
- De Vuyst, S., Bruneel, H., & Fiems, D. (2014). Computationally efficient evaluation of appointment schedules in health care. *European Journal of Operational Research*, 237(3), 1142-1154.
- Deceuninck, M., Fiems, D., & De Vuyst, S. (2018). Outpatient scheduling with unpunctual patients and no-shows. *European Journal of Operational Research*, 265(1), 195-207.
- Denton, B., & Gupta, D. (2003). A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11), 1003-1016.
- Erdogan, S. A., & Denton, B. (2013). Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing*, 25(1), 116-132.
- Erdogan, S. A., Gose, A., & Denton, B. T. (2015). Online appointment sequencing and scheduling. *IIE Transactions*, 47(11), 1267-1286.
- Fry, J. (1964). Appointments Systems in Hospitals and General Practice: Appointments in General Practice. *Journal of the Operational Research Society*, 15(3), 233-237.
- Gocgun, Y., & Puterman, M. L. (2014). Dynamic scheduling with due dates and time windows: an application to chemotherapy patient appointment booking. *Health Care Management Science*, 17(1), 60-76.
- Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9), 800-819.
- Hahn-Goldberg, S., Carter, M. W., Beck, J. C., Trudeau, M., Sousa, P., & Beattie, K. (2014). Dynamic optimization of chemotherapy outpatient scheduling with uncertainty. *Health Care Management Science*, 17(4), 379-392.
- Harper, P. R., & Gamlin, H. M. (2003). Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. *Or Spectrum*, 25(2), 207-222.

- Hong, T. S., Shang, P. P., Arumugam, M., & Yusuff, R. M. (2013). Use of simulation to solve outpatient clinic problems: a review of the literature. *South African Journal of Industrial Engineering*, 24(3), 27-47.
- Huang, Y. L. (2016). Appointment standardization evaluation in a primary care facility. *International Journal of Health Care Quality Assurance*.
- Huang, Y. L., & Bach, S. M. (2016). Appointment lead time policy development to improve patient access to care. *Applied Clinical Informatics*, 7(4), 954.
- Huang, Y. L., & Marcak, J. (2015). Grid patient appointment template design to improve scheduling effectiveness. *Journal of Healthcare Engineering*, 6(2), 239-258.
- Huang, Y., & Verduzco, S. (2015). Appointment template redesign in a women's health clinic using clinical constraints to improve service quality and efficiency. *Applied Clinical Informatics*, 6(02), 271-287.
- Jiang, R., Shen, S., & Zhang, Y. (2017). Integer programming approaches for appointment scheduling with random no-shows and service durations. *Operations Research*, 65(6), 1638-1656.
- Johnson, W. L., & Rosenfeld, L. S. (1968). Factors affecting waiting time in ambulatory care services. *Health Services Research*, 3(4), 286.
- Kim, S., & Giachetti, R. E. (2006). A stochastic mathematical appointment overbooking model for healthcare providers to improve profits. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(6), 1211-1219.
- Klassen, K. J., & Yoogalingam, R. (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, 18(4), 447-458.
- Klassen, K. J., & Yoogalingam, R. (2013). Appointment system design with interruptions and physician lateness. *International Journal of Operations & Production Management*, 33(4), 394-414.
- Klassen, K. J., & Yoogalingam, R. (2014). Strategies for appointment policy design with patient unpunctuality. *Decision Sciences*, 45(5), 881-911.
- Kong, Q., Lee, C. Y., Teo, C. P., & Zheng, Z. (2013). Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3), 711-726.
- Kortbeek, N., Zonderland, M. E., Braaksma, A., Vliegen, I. M., Boucherie, R. J., Litvak, N., & Hans, E. W. (2014). Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals. *Performance Evaluation*, 80, 5-26.
- Kuiper, A., & Mandjes, M. (2015). Appointment scheduling in tandem-type service systems. *Omega*, 57, 145-156.
- Kuiper, A., Kemper, B., & Mandjes, M. (2015). A computational approach to optimized appointment scheduling. *Queueing Systems*, 79(1), 5-36.
- LaGanga, L. R., & Lawrence, S. R. (2012). Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management*, 21(5), 874-888.
- Lee, S. J., Heim, G. R., Sriskandarajah, C., & Zhu, Y. (2018). Outpatient Appointment Block Scheduling Under Patient Heterogeneity and Patient No-Shows. *Production and Operations Management*, 27(1), 28-48.

- Leung, J. Y. (Ed.). (2004). Handbook of scheduling: algorithms, models, and performance analysis. CRC press.
- Lin, C. K. Y. (2015). An adaptive scheduling heuristic with memory for the block appointment system of an outpatient specialty clinic. *International Journal of Production Research*, 53(24), 7488-7516.
- Lin, C. K. Y., Ling, T. W. C., & Yeung, W. K. (2017). Resource Allocation and Outpatient Appointment Scheduling Using Simulation Optimization. *Journal of Healthcare Engineering*.
- Lindley, D. V. (1952, April). The theory of queues with a single server. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 48, No. 2, pp. 277-289). Cambridge University Press.
- Luo, J., Kulkarni, V. G., & Ziya, S. (2012). Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management*, 14(4), 670-684.
- Mak, H. Y., Rong, Y., & Zhang, J. (2014). Sequencing appointments for service systems using inventory approximations. *Manufacturing & Service Operations Management*, 16(2), 251-262.
- Mak, H. Y., Rong, Y., & Zhang, J. (2015). Appointment scheduling with limited distributional information. *Management Science*, 61(2), 316-334.
- Mancilla, C., & Storer, R. (2012). A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions*, 44(8), 655-670.
- Morikawa, K., & Takahashi, K. (2017). Scheduling appointments for walk-ins. *International Journal of Production Economics*, 190, 60-66.
- Morikawa, K., Takahashi, K., & Hirotsu, D. (2018). Performance evaluation of candidate appointment schedules using clearing functions. *Journal of Intelligent Manufacturing*, 29(3), 509-518.
- Nguyen, T. B. T., Sivakumar, A. I., & Graves, S. C. (2017). Scheduling rules to achieve lead-time targets in outpatient appointment systems. *Health Care Management Science*, 20(4), 578-589.
- Oleskovicz, M., Oliva, F. L., Grisi, C. C. D. H., Lima, A. C., & Custódio, I. (2014). Técnica de overbooking no atendimento público ambulatorial em uma unidade do Sistema Único de Saúde. *Cadernos de Saúde Pública*, 30, 1009-1017.
- Parizi, M. S., & Ghate, A. (2016). Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking. *Computers & Operations Research*, 67, 90-101.
- Pérez, E., Ntamo, L., Malavé, C. O., Bailey, C., & McCormack, P. (2013). Stochastic online appointment scheduling of multi-step sequential procedures in nuclear medicine. *Health Care Management Science*, 16(4), 281-299.
- Qi, J. (2016). Mitigating delays and unfairness in appointment systems. *Management Science*, 63(2), 566-583.
- Riise, A., Mannino, C., & Lamorgese, L. (2016). Recursive logic-based Benders' decomposition for multi-mode outpatient scheduling. *European Journal of Operational Research*, 255(3), 719-728.

- Rockart, J. F., & Hofmann, P. B. (1969). Physician and patient behavior under different scheduling systems in a hospital outpatient department. *Medical Care*, 463-470.
- Safdar, K. A., Emrouznejad, A., & Dey, P. K. (2016). Assessing the Queuing Process Using Data Envelopment Analysis: an Application in Health Centres. *Journal of Medical Systems*, 40(1), 32.
- Salzarulo, P. A., Mahar, S., & Modi, S. (2016). Beyond patient classification: Using individual patient characteristics in appointment scheduling. *Production and Operations Management*, 25(6), 1056-1072.
- Saremi, A., Jula, P., ElMekkawy, T., & Wang, G. G. (2015). Bi-criteria appointment scheduling of patients with heterogeneous service sequences. *Expert Systems with Applications*, 42(8), 4029-4041.
- Schuetz, H. J., & Kolisch, R. (2013). Capacity allocation for demand of different customer-product-combinations with cancellations, no-shows, and overbooking when there is a sequential delivery of service. *Annals of Operations Research*, 206(1), 401-423.
- Sickinger, S., & Kolisch, R. (2009). The performance of a generalized Bailey–Welch rule for outpatient appointment scheduling under inpatient and emergency demand. *Health Care Management Science*, 12(4), 408.
- Su, S., & Shih, C. L. (2003). Managing a mixed-registration-type appointment system in outpatient clinics. *International Journal of Medical Informatics*, 70(1), 31-40.
- Tai, G., & Williams, P. (2012). Optimization of scheduling patient appointments in clinics using a novel modelling technique of patient arrival. *Computer Methods and Programs in Biomedicine*, 108(2), 467-476.
- Tai, G., & Williams, P. (2013). Quantitative analysis of investment allocation over various resources of health care systems by using views of product lines. *International Journal of Systems Science*, 44(11), 2057-2067.
- Tang, J., Yan, C., & Cao, P. (2014). Appointment scheduling algorithm considering routine and urgent patients. *Expert Systems with Applications*, 41(10), 4529-4541.
- Tsai, P. F. J., & Teng, G. Y. (2014). A stochastic appointment scheduling system on multiple resources with dynamic call-in sequence and patient no-shows for an outpatient clinic. *European Journal of Operational Research*, 239(2), 427-436.
- Turkcan, A., Zeng, B., Muthuraman, K., & Lawley, M. (2011). Sequential clinical scheduling with service criteria. *European Journal of Operational Research*, 214(3), 780-795.
- Vanden Bosch, P. M., Dietz, D. C., & Simeoni, J. R. (1999). Scheduling customer arrivals to a stochastic service system. *Naval Research Logistics (NRL)*, 46(5), 549-559.
- Vink, W., Kuiper, A., Kemper, B., & Bhulai, S. (2015). Optimal appointment scheduling in continuous time: The lag order approximation method. *European Journal of Operational Research*, 240(1), 213-219.
- Wang, J., Chen, Y. F., & Xu, M. (2018). Optimization and approximation methods for dynamic appointment scheduling with patient choices. *Computers & Operations Research*, 92, 65-76.
- Wang, P. P. (1993). Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics (NRL)*, 40(3), 345-360.
- Wang, W. Y., & Gupta, D. (2011). Adaptive appointment systems with patient preferences. *Manufacturing & Service Operations Management*, 13(3), 373-389.

- Welch, J. D., & Bailey, N. J. (1952). Appointment systems in hospital outpatient departments. *The Lancet*, 259(6718), 1105-1108.
- White, D. L., Froehle, C. M., & Klassen, K. J. (2011). The effect of integrated scheduling and capacity policies on clinical efficiency. *Production and Operations Management*, 20(3), 442-455.
- Xiao, G., Dong, M., Li, J., & Sun, L. (2017). Scheduling routine and call-in clinical appointments with revisits. *International Journal of Production Research*, 55(6), 1767-1779.
- Yan, C., Tang, J., & Jiang, B. (2014). Sequential appointment scheduling considering walk-in patients. *Mathematical Problems in Engineering*, 2014.
- Yan, C., Tang, J., Jiang, B., & Fung, R. Y. (2015). Sequential appointment scheduling considering patient choice and service fairness. *International Journal of Production Research*, 53(24), 7376-7395.
- Zacharias, C., & Pinedo, M. (2014). Appointment Scheduling with No-Shows and Overbooking. *Production and Operations Management*, 23(5), 788-801.
- Zhang, Y., Shen, S., & Erdogan, S. A. (2017). Distributionally robust appointment scheduling with moment-based ambiguity set. *Operations Research Letters*, 45(2), 139-144.