



“十三五”国家重点图书出版规划项目

排序与调度丛书 (二期)

医院服务管理优化

——建模方法与决策

羊英 钟力炜 陈童 编著

清华大学出版社
北京

内 容 简 介

本书主要针对当前医院服务管理中的关键问题,整理编辑近年来在该领域的重要研究成果。全书包含8章,讨论了门诊管理与优化、急诊管理与优化、住院管理与优化、手术资源管理与优化、医技检查管理与优化、医院物料管理与优化及医院服务管理智慧化转型等问题,并对相应问题进行建模和求解,采用实际医院运营问题或数据作为算例分析。本书适合的读者包括:医疗运营管理研究人员、医院服务管理人员、管理科学与工程研究生和对医疗服务资源调度优化有兴趣的研究人员等。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

医院服务管理优化:建模方法与决策/羊英,钟力炜,
陈童编著. --北京:清华大学出版社,2024.11.--(排序
与调度丛书). --ISBN 978-7-302-67608-9

I. R197.32

中国国家版本馆 CIP 数据核字第 20245EA205 号

责任编辑:佟丽霞

封面设计:常雪影

责任校对:欧 洋

责任印制:沈 露

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市龙大印装有限公司

经 销:全国新华书店

开 本:170mm×240mm 印 张:18.5 字 数:340千字

版 次:2024年11月第1版 印 次:2024年11月第1次印刷

定 价:129.00元

产品编号:098295-01

《排序与调度丛书》编辑委员会

主 编

唐国春(上海第二工业大学)

副 主 编

万国华(上海交通大学)

沈吟东(华中科技大学)

吴贤毅(华东师范大学)

顾 问(按姓氏拼音排序,中英文分开排序)

韩继业(中国科学院数学与系统科学研究院)

林治勋(郑州大学)

秦裕瑗(武汉科技大学)

涂萃生(南开大学)

越民义(中国科学院数学与系统科学研究院)

Bo Chen(陈礴)(英国华威大学)

T. C. Edwin Cheng(郑大昭)(香港理工大学)

Nicholas G. Hall(美国俄亥俄州立大学)

Chung-Yee Lee(李忠义)(香港科技大学)

Michael Pinedo(美国纽约大学)

编 委(按姓氏拼音排序)

车阿大(西北工业大学)

陈志龙(美国马里兰大学)

高 亮(华中科技大学)

黄四民(清华大学)

李荣珩(湖南师范大学)

刘朝晖(华东理工大学)

谈之奕(浙江大学)

唐加福(东北财经大学)

唐立新(东北大学)

王 冰(上海大学)

王军强(西北工业大学)

张 峰(上海第二工业大学)

张玉忠(曲阜师范大学)

周支立(西安交通大学)

丛书序言

我知道排序问题是从 20 世纪 50 年代出版的一本名为 *Operations Research* (《运筹学》,可能是 1957 年出版)的书开始的。书中讲到了 S. M. 约翰逊(S. M. Johnson)的同顺序两台机器的排序问题并给出了解法。约翰逊的这一结果给我留下了深刻的印象。第一,这个问题是从实际生活中来的。第二,这个问题有一定的难度,约翰逊给出了完整的解答。第三,这个问题显然包含着许多可能的推广,因此蕴含了广阔的前景。在 1960 年前后,我在《英国运筹学》(*Operational Research*, 季刊,从 1978 年(第 29 卷)起改称 *Journal of the Operational Research Society*,并改为月刊)(当时这是一份带有科普性质的刊物)上看到一篇文章,内容谈到三台机器的排序问题,但只涉及四个工件如何排序。这篇文章虽然很简单,但我也从中受到一些启发。我写了一篇讲稿,在中国科学院数学研究所里做了一次通俗报告。之后我就到安徽参加“四清”工作,不意所里将这份报告打印出来并寄了几份给我,我寄了一份给华罗庚教授,他对这方面的研究给予很大的支持。这是 20 世纪 60 年代前期的事,接下来便开始了“文化大革命”,倏忽十年。20 世纪 70 年代初我从“五七”干校回京,发现国外学者在排序问题方面已做了不少工作,并曾在 1966 年开了一次国际排序问题会议,出版了一本论文集 *Theory of Scheduling* (《排序理论》)。我与韩继业教授做了一些工作,也算得上是排序问题在我国的一个开始。想不到在秦裕琰、林诒勋、唐国春以及许多教授的努力下,跟随着国际的潮流,排序问题的理论和应用在我国得到了如此蓬勃的发展,真是可喜可贺!

众所周知,在计算机如此普及的今天,一门数学分支的发展必须与生产实际相结合,才称得上走上了健康的道路。一种复杂的工具从设计到生产,一项巨大复杂的工程从开始施工到完工后的处理,无不牵涉排序问题。因此,我认为排序理论的发展是没有止境的。我很少看小说,但近来我对一本名叫《约翰·克里斯托夫》的作品很感兴趣。这是罗曼·罗兰写的一本名著,实际上它是以贝多芬为背景的一本传记体小说。这里面提到贝多芬的祖父和父亲都是宫廷乐队指挥,当贝多芬的父亲发现他在音乐方面是个天才的时候,便想将他培养成一名优秀的钢琴师,让他到各地去表演,可以名利双收,所以强迫他勤

学苦练。但贝多芬非常反感,他认为这样的作品显示不出人的气质。由于贝多芬有如此的感受,他才能谱出如《英雄交响曲》《第九交响曲》等深具人性的伟大乐章。我想数学也是一样,只有在人类生产中体现它的威力的时候,才能显示出数学这门学科的光辉,也才能显示出作为一名数学家的骄傲。

任何一门学科,尤其是一门与生产实际有密切联系的学科,在其发展初期那些引发它成长的问题必然是相互分离的,甚至是互不相干的。但只要研究继续向前发展,一些问题便会综合趋于统一,处理问题的方法也会与日俱增、深入细致,可谓根深叶茂,蔚然成林。我们这套丛书已有数册正在撰写之中,主题纷呈,蔚为壮观。相信在不久以后会有不少新的著作出现,使我们的学科呈现一片欣欣向荣、繁花似锦的局面,则是鄙人所厚望于诸君者矣。

越氏义

中国科学院数学与系统科学研究院

2019年4月

前 言

虽然近年来我国医疗技术和医疗资源不断发展,但相对我国超过 14 亿的庞大人口群体,如何提高医疗资源利用率,提高医疗服务质量,仍然是一个非常重要的课题。

医院服务管理虽然属于服务管理领域,但与一般服务行业管理存在较大差异。与一般企业运营管理不同,获利不是医院的首要目标,公益和社会责任对于我国的公立医院是更重要的目标。如何利用现有的人才、资金、技术和信息等资源为更多的患者服务,提高整体的医疗服务水平,是每个医院亟需解决的问题。

近年来关于医疗服务管理领域的定量研究成果越来越多,对促进医疗机构科学服务管理非常具有参考价值。本书围绕医院服务管理的几大关键问题进行研究,集结了近年来相关领域一些有代表性的成果,结合团队研究成果,进行编撰阐述,希望能为医院管理者和本领域有兴趣的学者提供参考。

全书分为 8 章。第 1 章对医院服务管理基本概念和内容进行介绍,梳理了该领域的关键问题并列出了本书的框架。第 2~7 章针对医院服务管理中关键问题进行分析:第 2 章介绍门诊管理与优化问题;第 3 章介绍急诊管理与优化问题;第 4 章介绍住院管理与优化问题;第 5 章着重分析手术资源管理与优化问题;第 6 章讨论医技检查管理与优化问题;第 7 章介绍医院物料管理与优化问题。第 8 章对医院服务管理智慧化转型进行了展望。

以下对本书体例进行说明:

(1)书中部分图表是从引用文献中摘录的,本书对图表中的英文进行了翻译,便于读者阅读。

(2)每章最后是本章文中引用的文献,文献按照在文中出现的顺序罗列。

(3)每个问题阐述方式都按照以下框架进行:相关研究概述—问题描述—模型建立与求解—管理启示。为了便于医院管理者和对医院服务管理感兴趣的读者理解本书内容,编写时省略了部分复杂算法及其数学推理过程,侧重于模型、算法和应用场景介绍,并对管理启示进行初步探讨。

本书在编写过程中参考了医疗管理研究领域诸多学者的成果,征询部分有

代表性文献的作者意见并获得作者的支持,引用编入这些成果,在此向万国华、唐加福、王冰、常健、杜奕、林慧丹、高更君、王治国、耿娜、项薇、刘洪伟、梁峰、梅勋等各位学者表示感谢!

本书的编写得益于唐国春教授的极力推动,在确定方向、书稿梳理、修改等环节都得到唐老师的无私帮助,特此表示感谢。在书稿完成后,感谢中国运筹学学会排序与调度分会的专家们给出了许多中肯的意见,对书稿修改提供了非常有价值的建议。感谢本书审稿人杜刚教授,对全书内容进行了细致审阅并提出了许多中肯的建议。在编撰过程中,上海第二工业大学的同事郝皓教授、何成教授、尹诗老师和范静老师等,上海市中医医院的王慧新、沈婷和张树瑛医生,上海市第一人民医院的许翔医生给予了许多帮助,感谢上海市第一人民医院郭万茹、陈秀秀和林碧云帮助整理和排版书稿。

由于编者学术水平以及写作时间限制,本书仍然存在诸多不足,敬请读者和同仁批评指正。

羊 英 钟力炜 陈 童

2023年3月

目 录

第 1 章 绪论	1
1.1 医院服务管理概述	1
1.1.1 医院服务管理概念与特征	1
1.1.2 医院服务管理的重要性	3
1.2 医院服务管理关键问题	3
1.2.1 医院服务管理的概念和属性	3
1.2.2 医院服务管理的范畴	4
1.2.3 关键问题梳理	6
1.3 医院服务管理的研究方法	10
1.4 本书内容框架	13
参考文献	15
第 2 章 门诊管理与优化	17
2.1 门诊管理概述	17
2.1.1 门诊管理基本内容	17
2.1.2 门诊管理流程	18
2.1.3 门诊管理关键问题	18
2.2 专家门诊预约管理问题	20
2.2.1 相关研究概述	20
2.2.2 问题描述与分析	21
2.2.3 模型建立	22
2.2.4 算例	29
2.2.5 管理启示	37
2.3 考虑患者行为的门诊资源配置	38
2.3.1 相关研究概述	38
2.3.2 问题描述与分析	39
2.3.3 模型建立	42

2.3.4	算例	45
2.3.5	管理启示	46
2.4	门诊服务调度问题	47
2.4.1	相关研究概述	47
2.4.2	问题描述与分析	49
2.4.3	基于 MDP 的在线调度方法	50
2.4.4	基于多目标优化的调度方法	54
2.4.5	管理启示	59
2.5	小结	61
	参考文献	62
第 3 章	急诊管理与优化	65
3.1	急诊管理概述	65
3.1.1	急诊管理基本内容	65
3.1.2	急诊管理流程	65
3.1.3	急诊管理关键问题	67
3.2	急诊患者调度问题	69
3.2.1	相关研究概述	69
3.2.2	问题分析与描述	69
3.2.3	基于仿真的调度方法	70
3.2.4	基于 MDP 的调度方法	73
3.2.5	管理启示	76
3.3	急诊手术调度问题	77
3.3.1	相关研究概述	77
3.3.2	问题分析与描述	78
3.3.3	预测反应调度方法	79
3.3.4	突发事件下急诊手术应急调度模型	86
3.3.5	管理启示	89
3.4	小结	90
	参考文献	91
第 4 章	住院管理与优化	93
4.1	住院管理概述	93
4.1.1	住院诊疗管理	93
4.1.2	住院服务管理	94

4.2	床位管理问题	95
4.2.1	相关研究概述	95
4.2.2	问题描述与分析	96
4.2.3	考虑 CMI 指数的床位分配	97
4.2.4	管理启示	103
4.3	护士排班问题	104
4.3.1	相关研究概述	104
4.3.2	问题描述与分析	105
4.3.3	考虑加班策略的护士排班问题	106
4.3.4	考虑护士偏好和公平的排班问题	113
4.3.5	管理启示	120
4.4	小结	120
	参考文献	121
第 5 章	手术资源管理与优化	123
5.1	手术资源管理理论与方法	123
5.1.1	手术资源管理运行机制	123
5.1.2	手术资源类型与管理	126
5.2	两阶段手术排程方法	128
5.2.1	相关研究概述	128
5.2.2	问题描述与分析	129
5.2.3	模型建立	130
5.2.4	应用案例	134
5.2.5	管理启示	140
5.3	基于博弈论的手术排序方法	141
5.3.1	相关研究概述	141
5.3.2	问题描述与分析	142
5.3.3	模型建立	142
5.3.4	算例	145
5.3.5	管理启示	148
5.4	基于匹配理论的手术排序方法	148
5.4.1	相关研究概述	148
5.4.2	问题描述与分析	149
5.4.3	模型建立	152

5.4.4	实施步骤及算例	153
5.4.5	管理启示	155
5.5	小结	156
	参考文献	156
第6章	医技检查管理与优化	160
6.1	医技检查管理概述	160
6.1.1	医技检查管理内容	160
6.1.2	相关研究现状	162
6.2	检查患者预约调度优化问题	165
6.2.1	问题描述与符号说明	165
6.2.2	模型建立	167
6.2.3	算例	172
6.2.4	管理启示	172
6.3	医技检查需求预测问题	174
6.3.1	问题描述与分析	174
6.3.2	模型与算法	175
6.3.3	算例	177
6.3.4	管理启示	179
6.4	小结	180
	参考文献	180
第7章	医院物料管理与优化	182
7.1	医院物料管理基本理论与问题	182
7.1.1	医院物料管理基本理论	182
7.1.2	医院物料管理问题	188
7.2	基于供给—分拆加工—配送的耗材管理	190
7.2.1	基于供给—分拆加工—配送的耗材管理调度研究	190
7.2.2	基于供给—分拆加工—配送的耗材配送路径优化	200
7.2.3	管理启示	210
7.3	基于数据挖掘的耗材管理	212
7.3.1	相关研究概述	212
7.3.2	问题描述与分析	213
7.3.3	模型建立	215
7.3.4	应用案例	218

7.3.5 管理启示	224
7.4 医疗废弃物回收管理	225
7.4.1 相关研究概述	225
7.4.2 问题描述与分析	226
7.4.3 模型建立	227
7.4.4 算例	233
7.4.5 管理启示	240
7.5 小结	241
附录 Canopy 算法程序	241
参考文献	244
第 8 章 医院服务管理智慧化转型	249
8.1 智慧医疗基本内容	249
8.1.1 智慧医疗基本理念	249
8.1.2 智慧医疗建设内容	252
8.1.3 智慧医疗发展现状与趋势	256
8.2 医院服务管理智慧化实现路径	260
8.2.1 实现路径	260
8.2.2 从信息化到智慧化	263
8.2.3 管理模式的转型	268
8.3 医院服务管理智慧化展望	269
8.4 小结	271
参考文献	272
附录 A 英汉排序与调度词汇	273

第1章 绪 论

1.1 医院服务管理概述

1.1.1 医院服务管理概念与特征

根据国家卫生健康委员会和国家中医药管理局 2020 年发布的《关于加强公立医院运营管理的指导意见》(国卫财务发〔2020〕27 号)(以下简称《意见》): 公立医院运营管理(hospital operations management)是以全面预算管理和业务流程管理为核心,以全成本管理和绩效管理为工具,对医院内部运营各环节的设计、计划、组织、实施、控制和评价等管理活动的总称,是对医院人、财、物、技术等核心资源进行科学配置、精细管理和有效使用的一系列管理手段和方法。《意见》指出公立医院运营管理要坚持五条基本原则,即公益性原则、整体性原则、融合性原则、成本效率原则和适应性原则。

医院服务是指在医院的医疗服务过程中,以物质资源为基础,为提高患者效用所进行的有创造价值的活动。医疗服务是指医疗服务机构(主体是医院)以患者和一定社会人群为主要服务对象,以医学技术为基本的服务手段,提供能满足人们需要的医疗保健服务,从而产生实际利益的医疗产出和非物质形态的服务。医疗是医院的中心工作,医疗服务管理是医院运营管理的核心内容^[1]。因此,医疗服务管理(medical service management)是公立医院运营管理的核心职能,是指在医院医疗系统活动全过程中进行的组织、计划、协调和控制,使之经常处于应有状态,并对变化了的客观环境有较快的适应性,达到最佳医疗效果和医疗效率的目的^[1]。医院医疗服务管理涉及医院多个部门和环节,主要包括门诊、急诊和住院部三个体系,另外涉及医疗支持系统,有:手术室、医技科室和消毒供应中心等。

国外很早就有学者对医院服务管理进行研究,如 Margaret 等^[2]在 *Operations Research and Health Care: A Handbook of Methods and Applications* 一书中集结了多篇相关研究成果,包括:医疗运营管理、公共政策与经济分析、诊疗管理建模与应用。Corinne 等^[3]在 *Operations Management in Healthcare: Strategy and Practice* 一书中从医疗服务质量、医疗服务提供、医疗服务成本和医疗服务管理柔性等方面阐述医院运营管理主要内容。我国医疗服务管理和国外有较大差别:

一是我国公立医院遵循公益性原则。根据以上文件所述,医院服务管理要以公益性为前提,以满足人民群众健康需求为出发点和落脚点,实现社会效益和服务效能最大化。这个目标和国外医疗机构服务管理目标有较大区别。

二是我国人口基数大,医院少。根据国家统计局数据,第七次全国人口普查结果公布,全国人口共 1 443 497 378 人,其中大陆有 1 411 778 724 人。2022 年医院共有 36 976 家,医疗卫生机构床位数每万人 69.16 张,医疗卫生机构诊疗人次为 84.16 亿人次。同时,随着人们生活质量的提高,对医疗服务质量要求也越来越高;此外,随着我国老龄化进程加快,对医疗服务需求量也越来越大。所以医院不仅要提供足够的医疗服务,还要提高服务质量,这对医院服务管理是一个很大的挑战。

医疗资源不充足,还要基于公益性原则为超过 14 亿人口提供医疗服务,这要求医院提高各项资源的利用率,尤其是医疗服务资源的利用率。罗利等^[4]在《医疗服务资源调度优化理论、方法及应用》中,论述了门诊资源、医技资源、手术室资源和病床资源调度优化及具体方法。其研究成果为医院医疗服务管理效率的提高和具体实施方法提供了参考借鉴。

随着技术的发展和医院医疗服务管理模式不断创新,医院医疗服务管理呈现出以下趋势:

(1) 规模化发展向强调服务质量转变

随着医改的不断深化,我国公立医院普遍面临从粗放扩张向精细化转变的压力。公立医院要想实现转型,向现代医院管理制度迈进,就必须提高医院服务管理精细化水平^[5]。

(2) 质量与效率的协同发展

医院在提高服务质量的同时,要兼顾服务效率,即在资源有限的情况下使其发挥最大作用,如提升患者满意度、降低运营成本等。

(3) 传统工作模式向横式新生产方式发展

包含两层意思:一是建立以疾病为单元的新型医疗组织;二是以多学科诊疗(multi-disciplinary treatment,MDT)为例的跨学科联合诊疗模式。

(4) 信息技术支撑精细化管理

当前医院医疗服务管理更加追求资源合理利用,以达到精细化的管理,而精细化管理离不开信息技术的应用。一方面信息技术能实现基础信息的搜集、存储、传输和共享;另一方面,当前以信息技术为基础的智慧医院、数字化医院的建设将大大提升资源利用率,实现医院医疗服务精细化管理。

(5) 病种成本及费用控制

近年来,我国卫生总费用快速增长,近几年更是达到同期 GDP 增幅的近两倍,给政府及患者造成了沉重的经济负担,患者就诊费用不断攀升、公共卫生医

疗资源的公平性难以保障。定价有效性缺失和补偿机制扭曲是造成上述情况的主要原因。为了解决上述问题,国家医疗保障局将“按疾病诊断相关分组(DRGs)付费”作为深化医保支付方式改革的核心内容^[6]。

2020年突如其来的新冠疫情席卷全球,医院医疗服务管理又面临新的问题和更大的挑战,这使得医疗服务精细化管理变得更为迫切。

1.1.2 医院服务管理的重要性

医院服务管理优化的研究以医院医疗服务精细化管理为目标,采用定量方法寻求优化管理方法和模式,旨在提高医院医疗资源利用率,优化医疗服务管理流程。研究意义主要体现在以下几个方面:

(1) 有利于实现医院医疗服务精细化管理

医疗服务主要涉及门诊、急诊和住院部,以及提供辅助的医技科室、手术室等,这些部门在提供医疗服务时应形成一个有机的整体。以往的研究大多停留在某个部分的或是定性的研究上,如果要实现精细化管理,则需要借助如运筹学、管理学、信息技术、计算机技术等多学科综合方法,定量研究其中的服务管理问题,给出科学决策建议。研究如何优化其管理问题并提出有效的实施建议将极大地推动医院医疗服务精细化管理,提高管理效率和资源利用率。

(2) 有利于提高医疗服务质量水平

医疗服务不同于其他服务行业的产品,具有专业性特点,医疗服务质量关系到患者的健康和生命。目前大的人口基数和少的医院数给医院(尤其是公立三级甲等医院)带来了巨大的医疗服务压力,如何在面对庞大的就诊量时仍然能够提供高质量的服务,是每个医院面临的具有挑战性的课题。因此,基于建模与决策量化的研究优化医疗服务管理,有利于提高医疗服务整体质量水平。

(3) 有利于提升患者就医体验

一方面,随着我国经济发展加速,人们的生活水平不断提高,对医疗服务及就医体验的要求也不断提升。另一方面,随着我国老龄化加剧,老年人占比越来越高,对医疗服务的需求也越来越多,如何满足老年患者的需求并使他们就医更加便利,就医体验更高,也是医院医疗服务管理亟待解决的问题。医疗服务管理优化,不仅要优化管理流程,更要深入探讨患者的特性,分析不同类别患者的需求,为他(她)们提供个性化的服务。

1.2 医院服务管理关键问题

1.2.1 医院服务管理的概念和属性

医院服务具有三个特性:

一是满足患者效用。效用是患者从医疗服务中得到满足的程度,医院服务是满足患者效用的创造价值的活动。

二是服务产品属性。患者对医疗服务的需求,在满足其核心利益的基础上,还要追求心理和精神上的满足。

三是实物与服务关系。医疗服务是以药品、医疗设备为基础的服务产品。

医院服务管理是指对医疗服务提供过程进行计划、组织、协调、监督等,以提高管理效率,使得医疗服务资源利用率最大化,并实现精细化管理。

医院服务管理是完成医疗服务的主要手段,是影响整个医院管理水平的核心环节,具有以下基本属性^[1]:

(1)整体性:医院服务管理涉及医院各个部门和环节,主要包括门诊、急诊和住院部三个部门;此外,还有医疗支持系统,包括手术室、消毒供应中心和其他医技科室等。这些部门提供医疗服务、医疗平台支持或辅助诊疗服务,并相互配合构成一个复杂且严密的医疗服务体系,医院服务管理应将各个环节形成一个有机整体,从整体上提升服务管理水平。

(2)连续性:医疗服务是一个连续的过程,每一个环节对最终的质量都至关重要。相应的,医院服务管理也必须是一个连续的过程,确保每个环节都能提供良好的医疗服务。此外,医院服务管理的连续性还包含:采用稳定的医院管理方法;医院服务管理项目的持续改进。《关于加强公立医院运营管理的指导意见》中也提出:要将现代管理理念、方法和技术融入运营管理的各个领域、层级和环节,提升运营管理精细化水平。

(3)动态性:医疗服务体系所处的社会环境和医疗环境随时都在发生变化,管理方法也应适时调整和变革。当前社会信息化进程加速,医疗技术不断更新迭代,老龄人群不断增加,对医疗服务需求增长迅速,加大了医院服务管理的难度。这些内部和外部的因素都促使医院必须动态调整其管理模式和方法。

(4)目的性:我国医院管理的基本原则是公益性,因此医院服务管理的根本目标是治病救人,“以患者为中心”,这是基本前提,围绕这一基本前提,医院服务管理的目的在于实现医疗服务的“安全、有效、方便、价廉”。在目前有限的医疗资源并不能完全满足全社会所有医疗服务需求的情形下,医疗服务效率显得尤为重要,通过医院服务管理,以尽可能低的医疗服务成本来实现尽可能好的医疗服务效果、完成尽可能多的医疗服务任务。

1.2.2 医院服务管理的范畴

医院服务其中一个重要特性是其高接触性,在医疗服务过程中,医务人员与患者接触的程度比较高,并且患者及其家属要求参与医疗服务过程。根据与

患者接触程度分析,医院服务管理包含以下具体职能:

(1)门诊管理

门诊是直接接受患者进行诊断、治疗、预防保健和康复服务的场所,同时也是进行医学教育和临床科研,以提高医院科学技术水平和医务人员业务能力的重要阵地。门诊工作是面向社会的重要窗口,因为它是医院接触患者时间最早、人数最多、范围最广的部门,因此,门诊对于满足患者需求,完成医院社会责任,具有重要的意义。门诊涉及从患者挂号、候诊、就诊,到医院提供检诊分诊、诊断、检验、检查、注射、治疗、取药等环节,是一连串的由多个环节组成的流程,在这个过程中,任何一个环节的阻塞都可造成门诊流程的不畅,给患者带来不便。因此要求医院应用系统管理理论和方法,优化门诊管理流程,提供高质量服务。

(2)急诊管理

医院急诊科是急诊医疗体系中的重要组成部分,它是急诊医学医疗、教研的基地,也是医院医疗工作的第一线,更是医院面向社会的重要窗口。急诊管理即对急诊服务提供过程的计划、组织、监督、评价等开展的工作。急诊科工作具有“急”“忙”“多学科性”等特点,急诊管理必须保障急诊治疗活动所需的人、设备、物料、技术、信息等资源的及时配备和补给。

(3)住院管理

住院管理是指针对入院患者实施诊疗活动过程的计划、组织、监督、评价等开展的工作。具体包括:住院诊疗组织结构的设计、医疗质量的监控、医务人员实施诊疗活动行为规范、诊疗技术的应用管理、规划提高住院诊疗整体水平的目标管理等。

同时还包括患者在住院过程中的护理服务管理,应为患者提供准确、安全、及时、有效的服务,对护理工作过程和服务实施控制与改进。

(4)手术资源管理

手术是医生对患者进行手术诊断、治疗和抢救的重要手段,涉及多种资源,如手术室、医生、麻醉师、护士、设备、耗材等。手术资源管理就是要合理调度使用医院已有的资源,为患者提供最好的服务。

(5)医技检查管理

医技科室包括药剂科、检验科、医学影像科(放射科、超声诊断科、核医学科等)、病理科、输血科、消毒供应中心和营养科等,因此医技检查种类繁多。医技检查管理就是要对各项检验检查进行合理安排,提高医技设备和人员的工作效率,缩短患者等待时间,减少患者在各个医技科室之间的辗转次数。

(6)医院物料管理

患者在接受治疗过程中不可避免地要使用到各种药品、医用耗材等,医院

物料管理主要指在对医疗服务过程中使用的各种药品、医用耗材等进行合理采购、传递、使用和记录,对物料的使用进行精细化管理,保证医疗效果的同时,降低患者的医疗费用。

1.2.3 关键问题梳理

医院服务管理涉及的问题多且复杂,本书遴选了其中部分关键问题进行分析、梳理。

1. 门诊管理问题

门诊工作的对象是医疗服务的最大人群,门诊管理中有很多值得研究的问题。

1) 门诊预约问题

患者日常体验比较差的,一般聚焦在:门诊排队长,预约专家难,滞留医院时间过长,医院就诊时间过短,在专家诊室里就诊时间过短等。患者到医院门诊有两种模式:一种是预约,另一种是自由就诊。采用全面预约就诊的优点显而易见,但是最优的预约机制应该是什么?如何保障医院的利益,既充分利用医生出诊时间不产生过多的空余,又不过长延时?是否可以采用“超售”的模式?同样,不能在某一个时段内使得患者积聚过多。因此,讨论分时段预约的问题以及患者爽约的问题非常有价值。但是从患者角度来看,如何获得预约的席位(即俗称的“挂上号”)可能更重要。已经有一些精准预约的研究和实践,通过预问诊方式把患者的病情和医生的医学背景、临床方向甚至临床研究的方向进行匹配,这是从医患两方面来进行机制设计。

2) 门诊患者异质性问题

门诊接待的患者类型多样:有的是预约的,有的是未预约直接进入的;有的是慢性病不太紧急的,有的是紧急的;有的是年轻的,有的是年纪较大的。面对形式多样的异质患者,传统的固定的排队方式不太适用,典型的例子如:医院对于80岁以上的老人还有一些其他急症的对象,需要安排尽早就诊;门诊中有特殊需求的患者,需要马上开好检验检查医嘱以使其有时间在医生的出诊时间内能够回来复诊等。所以在门诊管理中,如何针对异质患者设计合理的排队机制是一个重要的问题。

3) 其他问题

门诊当中还有一些其他问题,如一个患者有多种疾病需要在多个科室挂号就诊,患者在每个不同专科的就医时间安排,以及在医院的滞留时间,在医院空间动线距离,甚至占用医院重要的物理空间资源如电梯等,都需要做研究。

2. 急诊管理问题

患者来急诊有多个途径:

第一类是比较严重的,一般通过120救护车送到医院。120救护车在接诊后会将一些患者相关信息传送到就诊医院。在送诊过程中要考虑的问题包括:患者的病情与就诊医院急救能力匹配关系问题;医院能级分配问题;医院急救学科排队问题;城市路线行驶规划问题。到了医院后,医院进行接诊,然后送进抢救室进行急救。急救过程中涉及的问题包括:抢救室空间资源分配问题;抢救设备、抢救人员、抢救药品,以及后续要进行抢救的一系列大型设备如计算机断层扫描(computed tomography, CT)、磁共振成像(magnetic resonance imaging, MRI)、数字减影血管造影(digital subtraction angiography, DSA)和手术室安排问题。部分患者经过抢救以后,需要进一步治疗,将涉及的问题包括:抢救室、急诊重症加强护理病房(intensive care unit, ICU)和急诊留观室之间过渡的床位安排,在这个过程中,既要考虑患者的病情和床位安排,又要考虑相应区域配备医生的能级,还要考虑区域布局的院内感染问题。

第二类急诊患者和前述120救护车送抵的比较类似,是家属、路人或者警察将需要急救的患者送到医院进行治疗。但在他们到达前,医院很难获得相应的患者信息,所以是在接诊以后才会知道患者相关病情,对于医院在短期内进行人员、设备、空间的调配,以及相应的时间问题,均有较大管理难度。

第三类患者病情相对比较轻,是自行来院。急诊预检分诊需要了解患者的主诉现病史、既往史以及包括脉搏、血压、体温、血氧饱和度等生命体征的集成数据。在医院的急诊高峰期间将遇到很大的就诊压力,往往会形成排队现象,甚至有患者滞留的时间超过规定的时限(如有些医院规定:必须在到院10min之内进行就诊)。患者在分诊以后,按照级别进行就医。黄区红区是进到抢救室就诊,绿区黄区是在急诊台面上进行就诊。按照急诊分诊分级规范,患者级别越高,就越应该在更短的时间进行就诊,当患者不断增多,将涉及急诊的排队机制。有研究提出基于患者的生命体征以及病情进行研判的插队机制,但这些研究很难得到实证。

患者急诊就诊后做检验检查时,会涉及排队问题,同时要考虑患者在离开医生的看护下,他/她全程生命体征监控问题。当前,部分医院尝试使用各类穿戴设备,这些设备会产生大量的数据。这些数据在后台集成后,可以通过算法模型进行跟踪管理,确保患者被医护人员全程监控,病情一旦突变,将通过定位和追踪功能,及时进行医疗急救工作。

急诊抢救室有急诊或者ICU医生在实施抢救。但一旦患者病症涉及多学科,就需要急诊医生进行预先研判。专科医生从专科病房过来需要时间,研判

得越早,专科医生到达的时间越早,能够开展多学科救治的时间就越早,对患者预后会起到十分重要的作用。因此,急诊医生及时准确的研判非常重要。

急诊后的随访机制,是指如果患者病症情况有所变化,应当随时到医院进行就诊。但是现在缺乏对急诊患者离院期间的远程看护。现在市面上逐渐出现一些硬件和软件,如开发一些应用软件,在智能手机上安装,对患者身体状况进行监测,将数据传回医院,一旦发生异常,医院便及时提醒患者或患者家属,使患者能够及时来院救治。

3. 住院管理问题

住院管理中首要问题是床位的安排,目前我国医院床位安排有多种模式。

比较传统的模式是科室固定病区床位的模式,它明显的优势就是医生很明确其医疗资源容量,可以通过院、科、组三级管理,以日常经验来进行床位资源分配。

另外一种模式是住院床位相对固定,边界模糊。因为季节性或者其他原因会导致部分床位空余,导致资源浪费。解决这种问题的方法有两种:第一种是在科室内或者病区内协调;第二种是由医院统一的住院服务调度中心来进行全院的统一床位配置,这样可以充分利用床位资源。如华西医院建立了一个很大的住院服务中心,对全院的床位进行调配使用。这种模式下存在两个问题:一是,医生对于患者选择的自主权很差,尤其是在临床研究特别受关注的情况下,收治患者的权利不在医生,这种处理方式对于临床研究的开展会产生重大影响;二是,这种模式下,床位使用率会增高,但值得注意的是医院并不能盲目追求床位使用率,而周转率可能更加重要。其实,医院不希望医生通过患者压床来提高使用率,这将使得其他患者无法获得住院治疗机会。

医院通常使用住院日这个指标,它既解决了住院床位的效率问题,又有效地解决了住院费用尤其是均次费用快速增长的问题。如果把平均住院日和使用率结合在一起看,能够有效地描述一家医院床位使用情况。但是如果只用平均住院日进行考核,会有很多做法歪曲其原有的意义,比如只收治病情轻的患者或者让患者反复办理住院,以增加周转和出院人数等。目前可以通过疾病诊断相关分组(diagnosis related groups, DRGs)、患者重复住院率等指标进行修正。

住院管理中另一个重要的问题是住院护理,其中护士排班问题比较突出。护士排班有很多约束条件,因为病区的护士一般由护士长、主班护士、带教老师以及专科护士、责任护士组成。既有白天主班安排,又有床位上的责任护士和三班倒的护士排班要求,是一种多种排班形式的组合。在排班中还要考虑护士

因为反复倒班,日夜时间变化,导致的精力疲惫以后容易产生护理差错的问题。这个问题在医生的夜间值班排班中也会发生。这是医疗质量安全上一个很重要的隐患,都需要通过研究来解决这些问题。

4. 手术资源管理问题

手术一般是有创操作,牵涉的资源多,流程复杂,所以手术排班成为手术资源管理最重要的问题。现在大多数医院实行的手术排程主要是固定手术日的方法,也就是说一个科室或者一个组,在每周有几天是固定的手术日,在这几天这个组或者科室主要以手术工作为主,相应地这个组或科室的患者手术必须排在这几天。组或科室的手术日一般1~2年才会调整一次,导致手术室在有的手术日使用率很高,有的手术日使用率很低,结束的时间很早。手术室作为医院重要的医疗资源,在医院运营管理中通常以提高它的利用率为主要目标。而在一台手术中,还需要医生、手术室麻醉师、手术护士、重要医疗设备等的配合,所以还要研究这些资源之间的关系。各类专业人员医生有自己的专业,重要的医疗设备也不可能配置过多,需要研究如何充分使用它们。同时,手术室又有大小、洁净等级不同的情况,所以不同的手术必须安排在不同的手术室进行,使得手术排程问题很复杂。此外,患者手术结束以后的转运动线问题也是需要研究的。

5. 医技检查管理问题

检验检查由相关医技科室完成,具有共同特点:①技术专业,相对独立;②为临床诊疗提供客观依据,同时也对临床工作有日趋增强的指导作用;③技术发展既高度综合又高度分化,新兴边缘科学不断出现;④服务方式由辅助检查逐渐加入治疗职能;⑤投入成本高;⑥对仪器设备的依赖性日趋增多;⑦多学科人才优化组合日趋完善,质量控制技术日趋完善^[1]。

现代医学发展到今天,我们更加依赖先进的检验检查。但是医院的检验检查设备、人员和空间资源是有限的。

假设一个患者既有抽血、大小便取样检验,又有核素检查,还有前列腺超声检查。这些检验检查中有一些前后关系:如核素检查应该放在最后,免得患者带着核素在医院里到处走;而前列腺超声检查前需要憋尿,但如果患者前面取了尿样,那就可能需要再花两个小时喝水憋尿。有些检验检查是需要空腹的,有些需要餐后,等等。前置条件、部门容量、患者排程、转运动线,还有检验流水线及其他重大设备的使用等一系列的问题值得研究。

6. 医院物料管理问题

医院物料包括药物、耗材、标本等多种品类。

第一是药物配发管理。药物治疗是一种主要治疗手段,大多数患者在门诊、急诊就诊的最后一个环节是取药。这就涉及药房的管理以及药物配送的问题。传统模式下,药品供应商把药品送到医院以后,医院的中心库再向它的一级库和二级库进行发放。这当中存在药品保存管理、药品保存空间管理、药品有效期管理、药品退货的逆向物流等问题。目前供给—分拆加工—配送(supply-processing-distribution,SPD)模式被广泛应用于医院物料供应链管理。医院建设一个中心库,通过这个中心库进行发药。在这种模式下,须关注以下问题:中心库的空间管理、人员管理、自动发药机里的药品配置及药品添加、药品从中心药房到各个护理单元之间配送路线、特殊药品的冷链管理等。

第二是各种物料传送方式。目前部分医院通过传输管道来运送药品、标本及其他物料,需要考虑运送的方式和运送频率管理。

第三是医疗耗材管理。医疗活动中大量使用各种医疗耗材,包括收费耗材和不收费耗材、高值耗材和低值耗材,需要考虑耗材的精细化管理、耗材配送使用等问题。

第四是医疗废弃物运送。治疗过程中产生的大量医疗废弃物不同于一般垃圾,部分是有污染的,因此这类废弃物的运送和处理都应该加以严格管理,如何科学、合理地处理和运输医疗废弃物也是医院服务管理中的重要环节。

7. 其他新问题

自新冠疫情以来,医院又出现了一些新问题。以往医院内感染风险比较高的是手术室中心供应室、胃肠镜室、口腔科、产房、发热门诊等;可是新冠疫情暴发后,发热门诊和隔离病房成为了高感染风险区域,因此,发热门诊与隔离病房的管理是每家医院平战结合的重点,发热门诊的布局管理、隔离病房的资源管理、发热门诊和隔离病房的医生排班、医疗设备的配置、患者流程的管理、患者安全的管理等都存在很多问题。一旦有疫情暴发,如何在最短时间内把医院资源(包括空间资源、人员资源和物质资源等)做重新调度,这也是值得研究的。

1.3 医院服务管理的研究方法

在中国,学者使用各种不同的研究方法来研究医院服务管理。一方面非医疗系统行业相关的学者研究得到的结果不太被认可,这是因为医院这个行业的特殊性。医院虽是一个服务系统,但它是提供特殊的专业服务——医疗服务的,所以很难用一般的服务科学的方法来研究它。另一方面卫生事业管理专业

的研究者大多运用公共政策的一般研究方法进行定性的研究,而定量的研究比较少,运用数学、运筹学与控制论等一些偏理工科的方法就更少了。

基于政策背景进行研究,指向和国家卫生政策一般是同向的。一些研究是基于部分政策调整需要进行前期论证或者布局而产生的。当一些政策需要调整更改的时候,已有的相关研究结果就显得说服力很差。对于医院一院多址而产生的床位规模扩张,医院特需病房比例提高等一系列的卫生政策导向,是基于新冠疫情在中国实践中产生的问题的新认识。这与既往的研究结果有很大偏离。由此可见,前期研究结果的部分结论是不符合当前中国实践的。这与基于政策导向的研究结果预设有关,也反证了仅仅使用这些方法进行研究是不够的,甚至可能不会有高质量的研究结果产生。

所以,基于国家社会经济发展水平、公共政策有明确导向的背景去做微观的研究,运用更基础、更科学的方法来精准研究医院运营管理中的问题,可以得到一些模型和算法并实际解决这些问题,或者设计一些新的机制运用推广到实践中,这是有价值的。

以下总结了一些应用于医院服务管理的理论和方法。

1. 排序论

排序论作为运筹学的一个分支,对提高效率、资源开放和配置、工程进展的安排以及经济运行等方面都能起到辅助科学决策的作用^[7]。在最优化理论和应用中把“scheduling”“排序”“调度”这三者视为含义完全相同^[8],所以有时候也会用“调度”一词进行替代。排序理论是组合优化学科的重要组成部分。排序问题是工业生产中一类带有普遍性的问题。对于大型的、复杂的工作,排序好坏对工程费用的大小影响很大^[7]。排序论近年来被应用在工程管理、经济管理等多个领域中,医院运营管理中也应用了排序理论,最典型的是手术排程问题。手术排程涉及手术室、执刀医师、麻醉师、护士、辅助人员、手术设备和耗材等多种资源的调配,手术排程的目的是使手术室和有关手术资源的利用率最大化。

手术排程的整数规划数学模型是非线性的,是 NP-困难的,难以使用传统的分支定界法、整数规划法等精确求解方法进行求解。国际上绝大多数手术排程使用混合整数规划的数学模型;个别有建立排序论模型的,但是求解还是转化为数学规划方法,算法复杂,计算量大。随着计算技术和算法研究的发展,禁忌搜索算法、模拟退火算法、粒子群算法等元启发式算法出现,能较好地解决这类问题,并获得较好的近似最优解。这也给医院手术排程提供了更好的解决方案。

如果把手术患者看成等待加工的“工件”，把实施手术的执刀医师、麻醉师、护士和手术设备等看成加工“工件”时需要的“机器”，那么手术排程就可以表述成“工件”需要多台“机器”同时加工的机器排序问题。Brucker^[9]研究过这类机器排序问题，这也是 NP-困难的问题。把手术安排次序的排程问题表示成多台机器的排序问题，按照排序论，其记号是 $P_m | MMJ | C_{\max}$ ，优化的目标是使最后一台手术完成的时间尽可能早，即使安排的手术全部尽早完成。

本书第 5 章将介绍排序理论应用于手术排程的相关研究成果。

2. 排队论

排队论(queueing theory)，又称为随机服务系统理论，是通过研究各种服务系统在排队等待现象中的概率特性，从而解决系统最优设计与最优控制的一门学科^[10]，是研究系统由于随机因素而出现排队现象的规律的一门学科，是运筹学的一个重要分支。它适用于各种服务系统，如公共服务系统、通信系统、运输系统等^[11]。排队论在医院运营管理中得到广泛应用，如在医院内科门诊排队系统中的应用，在医院动态心电图检查排队系统中的应用，在口腔中心检查排队系统中的应用和在眼科专家门诊排队系统中的应用。

随着现代管理理论的发展及提高服务质量需求的提升，排队心理学加入排队论并应用到实际问题中。在研究患者就医排队问题时，若能综合考虑排队心理因素将能更好地提高患者满意度。

本书第 2 章将介绍排队论在门诊预约管理中的应用。

3. 博弈论

博弈论(game theory)，又称为对策论、赛局理论等，既是现代数学的一个新分支，也是运筹学的一个重要学科。博弈论主要研究公式化了的激励结构间的相互作用，是研究具有斗争或竞争性质现象的数学理论和方法。博弈论研究游戏中的个体的预测行为和实际行为，并研究它们的优化策略。生物学家使用博弈论来理解和预测进化论的某些结果。

一般认为，博弈主要可以分为合作博弈和非合作博弈。博弈论被广泛应用于经济学各个领域，在医疗服务研究中得到广泛应用，如分级诊疗^[12]、隐私保护^[13]、医联体合作^[14]等方面。

本书第 5 章着重介绍博弈论在手术资源管理方面的应用。

匹配理论是经济学理论中的一个重要研究领域，属于合作博弈范畴，最早由 L. S. Shapley 和 A. E. Roth 两位学者提出，对于两个不同的主体，如何匹配

才是恰当的,L. S. Shapley 和 A. E. Roth 分别从稳定匹配的抽象理论和市场制度的实际设计两个角度,回答了这个问题,从而获得 2012 年诺贝尔经济学奖。在医院运营管理中,有许多匹配问题,比如专家门诊匹配预约、手术安排,以及分级诊疗体系中社区医院与上级医院的匹配等。因此研究匹配理论以及将研究成果应用于相关的医院管理必将提升医院服务质量,提高患者满意度。

本书第 2 章和第 5 章将着重介绍应用匹配理论解决医院医疗服务管理问题的研究。

4. 流程优化

流程优化是运营管理中重要的方法。在流程的设计和实施过程中,要对流程不断地改进,以期取得最佳的效果。对现有工作流程的梳理、完善和改进的过程,称为流程优化。与“流程优化”紧密相关的概念是“流程再造”,流程再造是由美国的 Michael Hammer 和 James Champy 提出的,在 20 世纪 90 年代达到了全盛的一种管理思想,是一种企业活动,内容为从根本上重新、彻底地分析与设计企业程序,并进行管理,以追求绩效,使企业快速成长。企业再造的重点在于选定对企业经营极为重要的几项企业程序加以重新规划,以提高运营效果。目的是在成本、品质、对外服务和时效上得到重大改进。

医院业务流程分为核心流程和辅助流程两类:医院营销、院前急救、门诊流程、急诊流程、住院流程、临床诊断、临床检验、临床检查、麻醉与手术、临床护理、药品管理、跟踪服务等,是医院的核心流程;内部管理的一些流程则是辅助流程。流程优化理论和方法被应用于门诊流程优化和人力资源管理等方面。

本书第 2 章介绍流程优化在预约管理中的应用。

另外,还有各种优化算法也在医院服务管理优化中发挥了重要作用,如粒子群算法、支持向量机等,在此不一一介绍。详见本书后续章节。

1.4 本书内容框架

本书主要针对医院服务中关键管理问题展开分析和讨论,第 2~7 章总结和阐述相关内容在定量化方面的研究成果,并分析其应用价值和管理启示,图 1-1 是本书核心内容框架。

本书研究内容分为 8 章,分别为:

第 1 章为绪论。阐述医院医疗服务管理优化的研究背景和研究意义,对医院服务管理相关概念进行介绍,并分析整理该领域的关键问题,介绍当前研究方法和研究现状,最后介绍本书的内容框架。

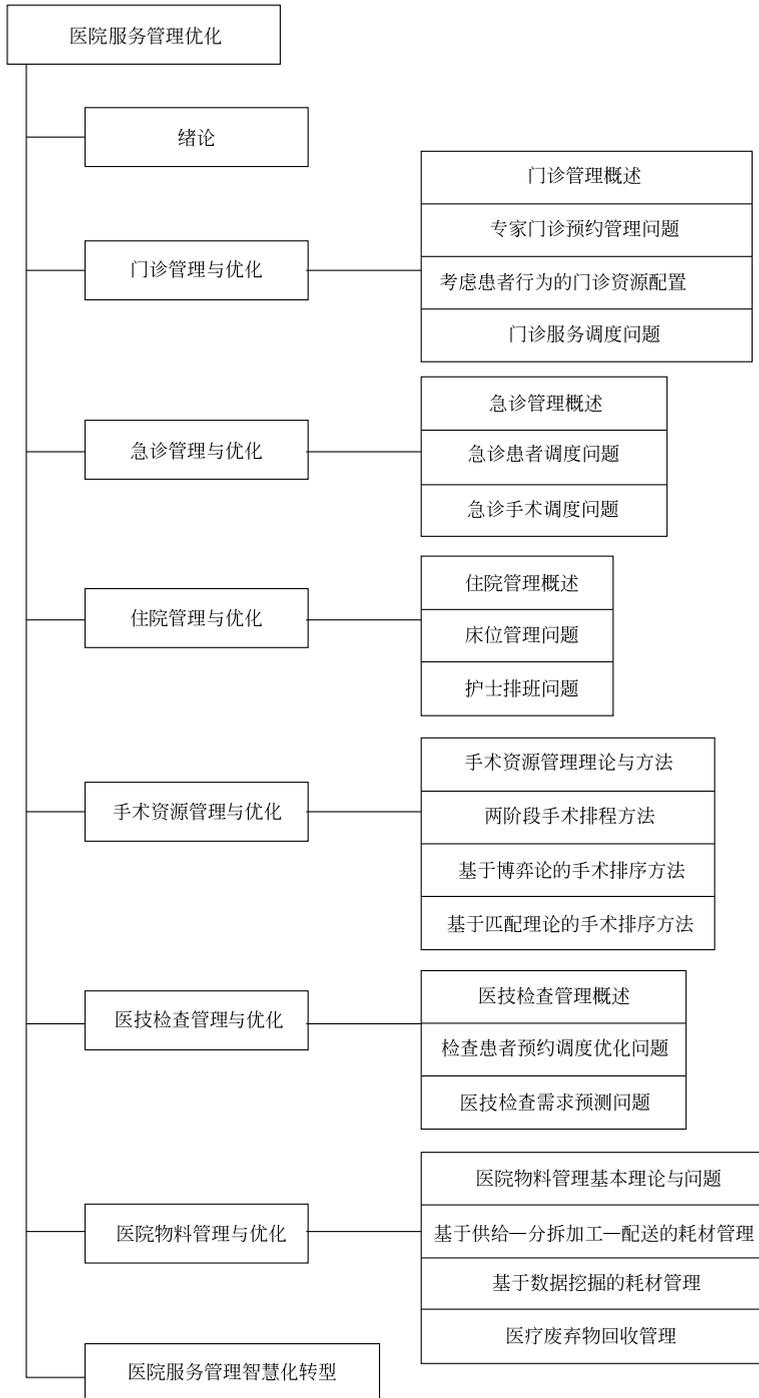


图 1-1 本书核心内容框架

第2章为门诊管理与优化。首先概述门诊管理基本内容和管理流程,分析该领域关键问题包括哪些;然后选择了三个主要问题进行研究,分别是:专家门诊预约管理问题,考虑患者行为的门诊资源配置,门诊服务调度问题。

第3章为急诊管理与优化。首先概述急诊管理的基本内容和流程,分析急诊管理中的关键问题,然后就急诊患者调度和急诊手术调度两个问题分别介绍优化方法。

第4章为住院管理与优化。首先对住院管理相关内容和基本问题进行阐述,然后针对床位管理问题和护士排班问题进行介绍。

第5章为手术资源管理与优化。首先介绍手术资源管理运行机制,然后对两阶段手术排程、基于博弈的手术排序和基于匹配理论的手术排序这三种方法分别进行介绍。

第6章为医技检查管理与优化。首先介绍医技检查管理基本内容和涵盖范围,然后对其中比较典型的检查患者预约调度优化问题和医技检查需求预测问题进行介绍。

第7章为医院物料管理与优化。首先对医院物料管理基本内容与方法进行介绍,然后基于供给—分拆加工—配送的耗材管理、基于数据挖掘的耗材管理及医疗废弃物回收管理等方面进行阐述。

第8章为智慧医院的医院服务管理发展趋势和展望。

参考文献

- [1] 张鹭鹭,王羽,薛迪,等. 医院管理学[M]. 北京:人民卫生出版社,2014.
- [2] MARGARET L B, FRANCOIS S, WILLIAM P P. Operations research and health care: a handbook of methods and applications[M]. New York: Kluwer Academic Publishers, 2005.
- [3] CORINNE M K, NANCY E D, MICHAEL R W. Operations management in healthcare: strategy and practice [M]. New York: Springer Publishing Company, 2016.
- [4] 罗利,石应康. 医疗服务资源调度优化理论、方法及应用[M]. 北京:科学出版社,2014.
- [5] 刘娟,王志粉,卢新翠,等. 构建精细化运营质量管理模式[J]. 中国卫生质量管理, 2020,27(2):1-3,25.
- [6] 刘雅娟. 基于 CCR 模型的病种成本管理体系构建及管理机制研究[J]. 中国医院管理, 2021,41(6):49-54.
- [7] 唐国春,张峰,罗守成,等. 现代排序论[M]. 上海:上海科学普及出版社,2003.
- [8] 唐国春. 排序论基本概念综述[J]. 重庆师范大学学报(自然科学版),2012,29(4):1-11.

-
- [9] BRUCKER P. Scheduling algorithms[M]. 3rd ed. Heidelberg:Springer, 2001.
- [10] 李军,徐玖. 运筹学[M]. 2版. 北京:科学出版社, 2003.
- [11] 周文正. 排队论模型在医疗服务系统中的应用研究[D]. 武汉:华中科技大学, 2011.
- [12] 杨力萌,梁峰,陈伟涛. 分级诊疗背景下双向转诊机制的演化博弈及仿真分析[J]. 工业工程与管理, 2021, 26(1):174-182.
- [13] 韩普,顾亮,张嘉明. 隐私保护视角下医疗数据共享意愿研究:基于三方演化博弈分析[J]. 现代情报, 2021, 41(3):148-158.
- [14] 张贝贝,陶帅,路伟,等. 博弈论视角下社会办医疗机构加入医疗联合体研究[J]. 中国医院管理, 2020, 40(10):36-38.

第2章 门诊管理与优化

门诊(outpatient)是直接接收患者并进行诊断、治疗和开展预防保健的场所,是医院和患者接触时间最早、人数最多的部门,是医院工作的重要组成部分。

2.1 门诊管理概述

门诊面对的患者多,病种最复杂,就诊时间最短,患者的期望值很高。特别是随着门诊功能逐步扩大,健康咨询、心理咨询、健康检查、康复指导等服务项目被普遍开展。医院门诊逐步向预防、医疗、保健、康复等多功能延伸,因此对门诊的管理和服务提出了新的更高的要求。也有部分医院使用现代门诊管理的原理和理念来运营,如五常法(常组织、常整顿、常清洁、常规范、常自律)、流程再造、全面质量管理(total quality management, TQM)、PDCA(即计划(plan)、执行(do)、检查(check)、处理(act))循环、公共关系管理(public relation management, PRM)等,可改善服务,塑造品牌,树立形象^[1]。

2.1.1 门诊管理基本内容

门诊是医院的门面,是医院的形象,也是医院管理中的重点部门。门诊管理主要针对门诊服务及门诊医疗等活动进行计划、组织、监督、协调等工作。

1. 门诊服务管理

门诊服务指为患者、患者家属及医护人员等在医院活动的人提供一个舒适的环境,保障门诊医疗活动正常开展。决定门诊服务质量的主要因素有:医院各部门空间布局的合理性、医院就诊流程安排的合理性、医院环境的整洁程度、医护人员的服务态度等。

当前,门诊患者就诊“三长一短”(即挂号、就诊、缴费排队时间长,看病时间短)的问题比较突出,为了解决这个问题,有部分医院提出使用门诊服务系数管理,即对门诊患者就诊有效服务时间的管理,其目的是分析患者排队的时间和原因,从而对症下药解决以上问题。

门诊服务系数是患者“接受服务时数”(T_s)与患者从进入到离开门诊所花费的总时数(T)之比。其公式为: $F(\text{门诊服务系数}) = T_s / T$ 。当前医院信息化进程加快,许多医院采用线上、线下多种挂号方式和缴费方式,大大缩短了患者挂号和缴费排队时间;许多医院采用预约制,患者到医院就诊等待时间也大大缩短了。门诊服务系数绝对值提高了,但是患者满意度不一定提高。因为在新的管理方式下,又涌现出新的问题:如预约制下许多患者对流程不熟悉,无法挂上号。因此如何从根本上提高门诊服务满意度,还需要对门诊挂号方式、号源分配、空间动线、流程安排等不断进行优化。

2. 门诊医疗管理

门诊医疗指为患者提供疾病诊治服务,区别于门诊服务管理,门诊医疗管理侧重于患者接受诊治的医疗过程的管理。决定门诊医疗质量的因素有:医护人员的专业能力、医疗过程中各环节安排的合理性、与诊疗活动相关部门的快速反应能力等。

以专家门诊的形式开展特需医疗服务是深化医院改革的一项重要措施,它有利于满足不同层次患者的需求,有利于医疗质量的提高,有利于医疗、教学、科研全面结合和发展,有利于调动高级技术职务人员的积极性,有利于体现和尊重知识价值和技术劳务价值,也有利于为患者提供更好的服务。但目前专家门诊中尚存在一些问题:一方面,好的专家号“一号难求”,导致许多真正需要专家诊治的患者见不到专家;另一方面,部分能够挂得到号的人可能只是普通疾病,不一定需要专家进行诊治。这种矛盾带来的问题显而易见:一是专家资源的浪费;二是专家无法尽其所长去救治最需要救治的患者。这种情况下,需要基于专家特长与偏好,将专家与患者进行匹配以解决问题。

门诊医疗过程除了到医生处就诊,还包括如图 2-1 所示的医生问诊后的检验/检查,再到返回医生处看结果、开药等过程。医疗过程安排合理性也将影响患者的就医体验和医院运作的效率。

2.1.2 门诊管理流程

图 2-1 所示是常见的门诊服务流程。每个患者进入门诊后都要经历几个主要环节:挂号、候诊、就诊、检验/检查、复诊/开药、付费、取药。医院门诊管理需要确保患者在整个流程中每个节点都得到满意的服务。

2.1.3 门诊管理关键问题

根据以上门诊服务流程以及 1.3 节的分析,梳理门诊管理关键问题。

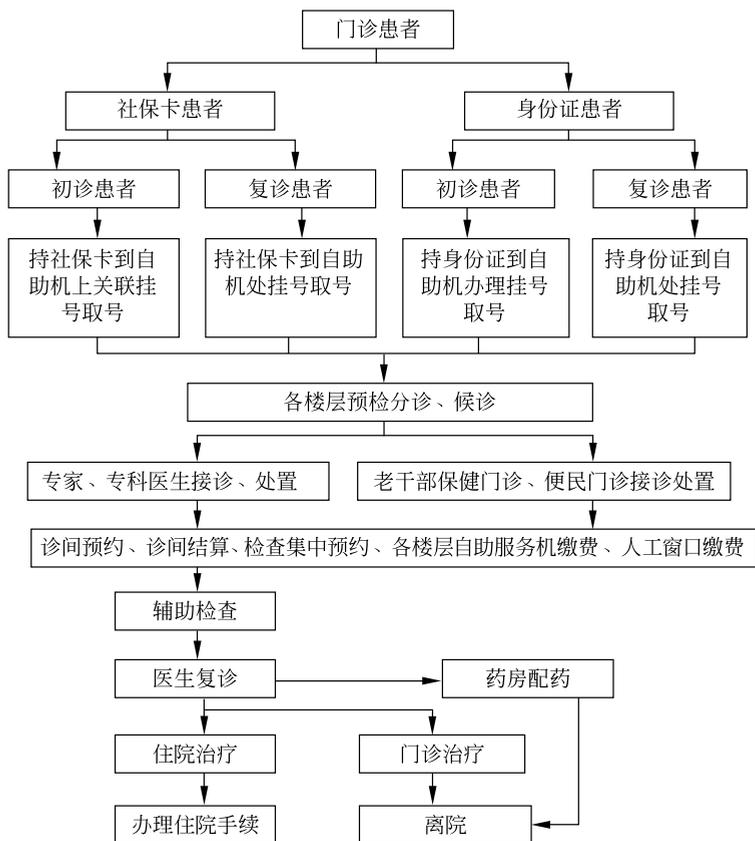


图 2-1 门诊服务流程

1. 专家门诊预约问题

专家是医院最重要的资源,要充分发挥其效能。目前存在的“挂号难”问题主要针对专家号难预约。如何使得最需要诊治的患者与最擅长诊治该患者病症的专家匹配上,不仅要考虑患者的需求,还需要在专家门诊预约流程、预约机制以及后台的算法上进行研究,进行精细化管理。

2. 患者就医行为模式研究

当前我国各大医院信息化程度不断提高,“互联网+”医院已经得到广泛应用,患者有多种方式获得诊疗服务。

同时,患者预约挂号也呈现多种方式:

其一是线上方式,可以通过类似于微医网(挂号网)网站,或者各种 App,各大医院的公众号等;

其二是线下方式,包括直接到人工窗口挂号、到医院自助机上挂号等方式。

患者在进行预约挂号时其行为模式有何特点?患者爽约的概率有多高?如果“超售”会带来什么样的影响?如果安排合理插队,患者是怎样的态度?在门诊机制设计中应当考虑患者就医行为模式,这也是当前门诊管理面临的重大课题。

3. 门诊服务调度

门诊资源包括人(各级医生、护士、护工、后勤人员等)、物(药品、耗材、清洁用品等)和信息(门诊患者挂号信息、诊疗信息、排队信息等),如何合理利用这些资源并使其效能最大化,需要对其进行合理调度。门诊管理需要合理配置这些资源,为患者提供高质量服务。由于患者到达具有随机性,且择医行为具有不确定性,因此以患者为中心的门诊服务调度是一个非常值得研究的课题。

4. 门诊流程优化问题

患者进入医院后,看诊过程包含哪些环节,各个环节间如何衔接?各个环节资源应如何配备?这些问题的解决将大大提高患者的体验和满意度。

本章后续将选择部分关键问题展开研究。

2.2 专家门诊预约管理问题

2.2.1 相关研究概述

中国是拥有超过 14 亿人口的大国,中国医疗保健系统面临巨大的压力。根据国家统计局数据,全国医院门急诊次数到 2022 年已达到 79.80 亿人次,全国医院数 36 976,诊疗 38.2 亿人次。中国居民医疗服务增加较快的主要因素是人口总量增加、人口年龄结构变化和患病率上升等。中国的医疗资源供需严重失衡(医疗资源总量不足、医疗资源分布不均衡、优质资源利用率不高),在医疗工作中重要的一环——门诊医疗中尤其突出。

全球对于门诊预约的研究始于 1952 年两篇开创性著作^[1-2]。Cayirli 和 Veral^[3]综述了 1952—2003 年对于门诊预约的定义、具有科室特色的最优预约规则、考虑多项因素(如患者失约、现场挂号等)的排队模型和引入公平性的预约系统评价等。Gupta 和 Denton^[4]指出门诊预约管理要为紧急患者预留时间,预约管理要提升门诊优质资源的利用率。

对于预约系统的优化,大多数的研究是采用排队论和收益管理的理论与方法。排队论优化模型的研究主要聚焦患者队列的形成、预约规则、到达过程、等待时间、医生空闲时间等^[1,5-7]。Moore 等^[8]、Gupta 和 Wang^[9]把收益管理理

论应用于医院门诊预约系统,还有学者把该理论应用于紧急患者的优先级确定等方面^[7,10-13]。

中国国内研究门诊预约主要从资源管理的角度,通过优化门诊资源配置、推算最合理的出诊医生数量,进而优化患者的等待时间与医院服务成本^[14-16]。然而对于中国医院现行的“单向预约、单向选择、先选先得”的门诊预约机制很少有人提出异议。2009年,北京市卫生局推出“层级预约制”,并在首都医科大学宣武医院等几家医院实施。此机制对高级专家门诊和普通专家门诊进行分类管理,患者就诊必须由普通专科门诊医生根据病情进行安排。这种预约规则模仿国外家庭医生制,但是还没有能够实现“全科医生首诊—专科医生预约”的就诊模式。

预约不能增加医疗资源的供给,但好的预约机制会让医患双方同时受益。2012年美国学者 Roth 和 Shapley 因在“稳定匹配理论和市场设计实践”领域中的杰出贡献荣获诺贝尔经济学奖。该研究成果基于合作博弈思想实现了对市场资源进行稳定、高效的匹配,一经提出便受到关注。到目前为止,已经被广泛地应用于住院医师分配问题、学生择校问题、器官移植匹配问题等多个领域。该研究成果是否也适用于专家与患者之间的匹配呢?

2.2.2 问题描述与分析

目前我国各医院实行“患者选择医生”的模式,但该模式存在一些不合理的

地方。门诊实行“患者选择医生”欠合理。2000年7月18日卫生部、国家中医药管理局联合制定了《关于实行病人选择医生,促进医疗机构内部改革的意见》(卫医发[2000]234号),把“患者选择医生”以文件的形式“规定”下来。多数患者出于对病情的担忧,在这个文件实施后,更愿意选择专家门诊。《关于印发2022年度全国三级公立医院绩效考核国家监测分析情况的通报》(国卫医政函[2024]30号)中提到:2022年全国三级公立医院的卫生技术人员中具有副高级职称及以上的医务人员比例为19.1%。根据国家统计局数据,2022年全国卫生技术人员总数为1165.79万人,其中执业医师人数仅有372.18万人。根据以上比例推算,具有副高级职称以上的执业医师(即患者认为的“专家”)人数相对全国超过14亿人口的就医需求还是相对紧缺的。因此,专家门诊号源紧张,供不应求,患者为了能得到心仪的专家诊治,往往需要通宵排队挂号或在线上长时间蹲守抢号。由于医学专业复杂,医患信息不对称,使患者的“自主择医”在大多数情况下成为了“盲目择医”,体现为:

(1)患者选择的医生,其专业与患者的病情无关(俗称“挂错号”)。医学不断发展,不断细化,使得不同医生所擅长的专业领域差别很大。患者由于缺乏

医学专业知识或是盲目追求“知名专家”，往往选错医生，既浪费医生的时间与精力，也耽误自身疾病的诊治。

(2)需要专家救治的疑难重症患者未能预约成功(俗称“挂不上号”)。为了让医生在有限的门诊时间内给予患者优质的诊疗服务，医院普遍实施就诊时间内固定医生就诊的人数。由于医生无法事先筛选患者的病情，“先选先得”的“规则”会“拒绝”那些疑难重症的患者使其不能及时得到医学专家的救治。

(3)医生无法诊治其擅长专业领域内的疾病。医生无权选择患者，诊治的对象往往不是擅长专业领域内的患者，既浪费了优质医疗资源，也阻碍了临床科研的发展。

“盲目就医”的现象在发达国家很少发生。在发达国家，患者由签约的家庭医生安排到专科医生处就诊。由于家庭医生对患者的病情和专科医生的专业特长都比较了解，因此很少发生患者“盲目就医”的现象。但是中国目前以及未来一段时间内还没有办法全面推行家庭医生制度。

针对以上问题，本部分以稳定匹配理论为基础，建立“患者选择专家”和“专家选择患者”的双向匹配的预约机制来优化门诊资源。

2.2.3 模型建立

1. 均衡匹配

基于双边匹配理论，本文提出专家门诊匹配预约，这个预约的核心是均衡匹配算法。

1) 一对一均衡稳定匹配

如果主体对 (m_i, w_j) 是一对一匹配 $\{M, W; P_M, P_W\}$ 的匹配对，记为 $(m_i, w_j) \in \mu$ ；如果主体对 (m_i, w_j) 是非匹配对(中意对或非中意对)，记为 $(m_i, w_j) \notin \mu$ 。

用0-1变量 x_{ij} 来表示主体对 (m_i, w_j) 的状态。如果主体对 (m_i, w_j) 是匹配对 $(m_i, w_j) \in \mu$ ，那么 $x_{ij} = 1$ ；否则，对于非匹配对 $(m_i, w_j) \notin \mu$ ，有 $x_{ij} = 0$ ， $i = 1, 2, \dots, n; j = 1, 2, \dots, n$ 。因此，主体集 $M = \{m_1, m_2, \dots, m_n\}$ 和主体集 $W = \{w_1, w_2, \dots, w_n\}$ 之间的一对一双边匹配 $\{M, W; P_M, P_W\}$ 是稳定匹配的充分必要条件为^[17]

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\ x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{cases} \quad (2.2.1)$$

稳定匹配的个数是非常多的^[17],如何在稳定匹配中寻找“较好”的匹配是非常有意义的工作。1990年Feder在博士论文^[18]中提出使得 $\max\left\{\sum_{(m_i, w_j) \in \mu} a_{ij}, \sum_{(m_i, w_j) \in \mu} b_{ij}\right\}$ 为最小的优化目标,并定义均衡稳定匹配(balanced stable matching)如下。

定义1 在一对一匹配问题 $\{M, W; P_M, P_W\}$ 的稳定匹配 μ 中,如果 μ^* 满足:

$$\max\left\{\sum_{(m_i, w_j) \in \mu^*} a_{ij}, \sum_{(m_i, w_j) \in \mu^*} b_{ij}\right\} = \min_{\mu} \max\left\{\sum_{(m_i, w_j) \in \mu} a_{ij}, \sum_{(m_i, w_j) \in \mu} b_{ij}\right\}$$

那么称 μ^* 是 $\{M, W; P_M, P_W\}$ 的均衡稳定匹配,简称均衡匹配。

由于 $\sum_{(m_i, w_j) \in \mu} a_{ij} + \sum_{(m_i, w_j) \in \mu} b_{ij} + \left| \sum_{(m_i, w_j) \in \mu} a_{ij} - \sum_{(m_i, w_j) \in \mu} b_{ij} \right| = 2 \max\left\{\sum_{(m_i, w_j) \in \mu} a_{ij}, \sum_{(m_i, w_j) \in \mu} b_{ij}\right\}$, 所以均衡匹配考虑两个优化目标:一是使主体双方总的序值之和 $\sum_{(m_i, w_j) \in \mu} a_{ij} + \sum_{(m_i, w_j) \in \mu} b_{ij}$ 为最小,称为平等稳定匹配(egalitarian stable matching)^[2],这个目标反映以双方主体总的“愿望值”尽可能高来选择;二是使得对两个主体总的序值之差 $\left| \sum_{(m_i, w_j) \in \mu} a_{ij} - \sum_{(m_i, w_j) \in \mu} b_{ij} \right|$ 为最小,称为性别平等稳定匹配(sex-equal stable matching)^[2],这个目标反映双方主体总的意愿相差较小,也就是对于双方主体是比较“均等”的。

由此得到 $\{M, W; P_M, P_W\}$ 的均衡匹配数学模型:

$$\min \max\left\{\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij}, \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_{ij}\right\} \quad (2.2.2)$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\ x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{cases}$$

上述模型的最优解就是 $\{M, W; P_M, P_W\}$ 的均衡匹配。

引入变量 U , 可将此模型转化为整数线性规划模型:

$$\begin{aligned}
 & \min_{x_{ij}} U \\
 & \left\{ \begin{aligned}
 & U \geq \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij} \\
 & U \geq \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_{ij} \\
 \text{s. t. } & \begin{cases}
 \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\
 \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\
 x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\
 x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n
 \end{cases}
 \end{aligned} \right.
 \end{aligned} \tag{2.2.3}$$

2) 一对多均衡稳定匹配

类似一对一匹配,提出一对多匹配的均衡稳定匹配。

主体集 $M = \{m_1, m_2, \dots, m_n\}$ 和主体集 $W = \{w_1, w_2, \dots, w_m\}$ 之间的一对多双边匹配 $\mu^{[2]}$ 是在 M 和 W 之间的一对多映射,即主体集 M 中的每一个主体 m_i 都有主体集 W 中的 $s(i)$ (其中 $s(i) \geq 1, i = 1, 2, \dots, n$) 个主体 $w_{j_1}, w_{j_2}, \dots, w_{j_{s(i)}}$ 作为其映像与之相对应,而且主体集 M 中所有主体的映像就是主体集 W , 即 $\sum_{i=1}^n s(i) = m$; 也可以表示为 $\mu(m_i) = \{w_{j_1}, w_{j_2}, \dots, w_{j_{s(i)}}\}$ 和 $\mu(w_{j_h}) = m_i, h = 1, 2, \dots, s(i)$, 或者表示为 $(m_i, w_{j_1}, w_{j_2}, \dots, w_{j_{s(i)}})$ 。此外,规定当 $m_i \neq m_k$ 时有 $\mu(m_i) \cap \mu(m_k) = \emptyset$ 。对于这样定义的一对多双边匹配 μ , 我们称 $(m_i, w_{j_1}), (m_i, w_{j_2}), \dots, (m_i, w_{j_{s(i)}})$ 是匹配 μ 的 $s(i)$ 个匹配对,称 m_i 关于匹配 μ 的映像为 $\{w_{j_1}, w_{j_2}, \dots, w_{j_{s(i)}}\}$, 称 w_{j_h} 关于 μ 的映像为 $m_i, h = 1, 2, \dots, s(i)$ 。如果 $P_M = [a_{ij}]_{n \times m}$ 是主体集 M 对主体集 W 的意向矩阵, $P_W = [b_{ij}]_{n \times m}$ 是主体集 W 对主体集 M 的意向矩阵,那么这个一对多匹配可以表示为 $\{M, W; P_M, P_W, S\}$, 其中向量 $S = (s(1), s(2), \dots, s(n))$ 。

对于在主体集 M 和主体集 W 中的主体 $m_i \in M$ 和主体 $w_j \in W$, 把 (m_i, w_j) 称为主体对。假定这个主体对不是 μ 的匹配对, 并且 m_i 和 w_j 关于匹配 μ 的映像分别是 $\{w_{k_1}, w_{k_2}, \dots, w_{k_{s(i)}}\}$ 和 m_l , 即 $\mu(m_i) = \{w_{k_1}, w_{k_2}, \dots, w_{k_{s(i)}}\}$, $\mu(w_j) = m_l$, 如果有

$$a_{ij} < \max\{a_{ik_1}, a_{ik_2}, \dots, a_{ik_{s(i)}}\} \text{ 和 } b_{ij} < b_{lj} \tag{2.2.4}$$

那么称 m_i 和 w_j 是**中意对**；如果有

$$a_{ij} \geq \max\{a_{ik_1}, a_{ik_2}, \dots, a_{ik_{s(i)}}\} \text{ 或者 } b_{ij} \geq b_{lj} \quad (2.2.5)$$

那么称 m_i 和 w_j 是**非中意对**。

因此,一对多双边匹配 μ 的主体集 M 和 W 之间所有的主体对 $(m_i, w_j) \in \mu$ 可分为三类:匹配对、中意对和非中意对。如果主体集 M 和 W 关于匹配 μ 不存在中意对时,称匹配 μ 是**稳定匹配**。因此,稳定匹配 μ 中除了匹配对,其余都是非中意对。

对 $\{M, W; P_M, P_W, S\}$ 的一个匹配 μ ,把 μ 的匹配对记为 $(m_i, w_j) \in \mu$,把 μ 的非匹配对(中意对或非中意对)记为 $(m_i, w_j) \notin \mu$ 。

定义 2 在 $\{M, W; P_M, P_W, S\}$ 的所有稳定匹配 μ 中,若 μ^* 满足: $\max\left\{\sum_{(m_i, w_j) \in \mu^*} a_{ij}, \sum_{(m_i, w_j) \in \mu^*} b_{ij}\right\} = \min_{\mu} \max\left\{\sum_{(m_i, w_j) \in \mu} a_{ij}, \sum_{(m_i, w_j) \in \mu} b_{ij}\right\}$, 则称 μ^* 是一对多匹配问题 $\{M, W; P_M, P_W, S\}$ 的**均衡匹配**。

3) 一对多均衡稳定匹配转换为一对一均衡稳定匹配

在实际的一对多匹配问题中,往往并不能保证两个主体集的主体恰好完全得到配对,即主体集 W 中主体的总数 $|W| = m$ 可能并不恰好等于主体集 M 中的所有主体可以匹配的元素之和 $\sum_{i=1}^n s(i)$ 。对此通过在主体集 W 或 M 中引进虚拟主体来达到两者的平衡。

如果 $m < \sum_{i=1}^n s(i)$,那么在主体集 W 中引进 $\sum_{i=1}^n s(i) - m$ 位虚拟主体 $w_{m+1}, w_{m+2}, \dots, w_{\sum_{i=1}^n s(i)}$,并假定主体集 M 中的主体 m_i 对虚拟主体 w_{m+j} 的意向序值 $a_{ij} = m+1, i=1, 2, \dots, n; j=m+1, m+2, \dots, \sum_{i=1}^n s(i)$ 。而虚拟主体 w_{m+j} 对 M 中的主体 m_i 的意向序值 $b_{ij} = 1, i=1, 2, \dots, n; j=m+1, m+2, \dots, \sum_{i=1}^n s(i)$ 。在一个匹配 μ 中,如果 $(m_i, w_{m+j}) \in \mu$,则表示 M 中的 m_i 匹配一个虚拟的主体,即 m_i 实际少了一个匹配对象。

如果 $m > \sum_{i=1}^n s(i)$,那么在主体集 M 中引进一个虚拟主体 m_{n+1} ,并假定 $s(n+1) = m - \sum_{i=1}^n s(i), a_{n+1,j} = j, b_{n+1,j} = n+1$,其中 $j=1, 2, \dots, m$ 。在一个匹配 μ 中,如果 $(m_{n+1}, w_j) \in \mu$,则表示 W 中的 w_j 匹配 M 中的虚拟主体 m_{n+1} ,即 w_j 在 μ 中实际上没有找到匹配对象。

为了在匹配运算时保证主体集 M 中各主体偏好序值取值范围的一致性,在保证原有矩阵顺序结构不变的基础上使 \mathbf{P}_M 的每一行之和均等于常数 $\sum_{j=1}^m j = \frac{m(m+1)}{2}$, 为此,对矩阵 \mathbf{P}_M 做如下的一致性处理。

不失一般性,假定主体集 M 中的主体 m_i 将主体集 W 中的 m 个主体分为 q_i 类,其中第 k 类主体的个数有 p_{ik} 个, $k=1,2,\dots,q_i$, $\sum_{k=1}^{q_i} p_{ik} = m$ 。主体集 M 中的主体 m_i 对主体集 W 中第 k 类主体的偏好序值设为 k 。为此,构造主体集 M

对主体集 W 的偏好矩阵 \mathbf{P}_M , 令: $a_{ij} = \frac{(\sum_{l=1}^{k-1} p_{il} + 1) + \sum_{l=1}^k p_{il}}{2}$ 。

上述定义的意向矩阵 \mathbf{P}_M 满足:

(1) 不会改变主体集 M 对主体集 W 的偏好顺序;

(2) \mathbf{P}_M 每一行的序值之和都为常数 $\sum_{j=1}^m j = \frac{m(m+1)}{2}$ 。

证明:

(1) 设 ω_j 属于第 k 类, $\omega_{j'}$ 属于第 k' 类, 且 $k' > k$ 。则有

$$\begin{aligned} a_{ij'} - a_{ij} &= \frac{(\sum_{l=1}^{k'-1} p_{il} + 1) + \sum_{l=1}^{k'} p_{il}}{2} - \frac{(\sum_{l=1}^{k-1} p_{il} + 1) + \sum_{l=1}^k p_{il}}{2} \\ &= \frac{\sum_{l=k}^{k'-1} p_{il} + \sum_{l=k+1}^{k'} p_{il}}{2} > 0 \end{aligned}$$

即对于矩阵 \mathbf{P}_M , M 中的 m_i 对 $\omega_{j'}$ 的偏好序值比对 ω_j 的偏好序值要大。

(2) 矩阵 \mathbf{P}_M 中第 i 行的偏好序值之和为

$$\begin{aligned} \sum_{j=1}^m a_{ij} &= \sum_{j=1}^m \frac{(\sum_{l=1}^{k-1} p_{il} + 1) + \sum_{l=1}^k p_{il}}{2} = \sum_{k=1}^{q_i} p_{ik} \frac{(\sum_{l=1}^{k-1} p_{il} + 1) + \sum_{l=1}^k p_{il}}{2} \\ &= \sum_{k=1}^{q_i} p_{ik} \frac{\sum_{h=1}^{p_{ik}} (\sum_{l=1}^{k-1} p_{il} + h)}{p_{ik}} = \sum_{j=1}^m j = \frac{m(m+1)}{2} \end{aligned}$$

同样,可以对 \mathbf{P}_W 进行一致性处理,在保证 \mathbf{P}_W 顺序结构不变的基础上使得

\mathbf{P}_W 的每一列的和均等于常数 $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ 。

因为主体集 M 中的每一个主体 m_i 都有主体集 W 中的 $s(i)$ (其中 $s(i) \geq 1$, $i=1, 2, \dots, n$) 个主体 $w_{j_1}, w_{j_2}, \dots, w_{j_{s(i)}}$ 作为其映像与之相对应, 为了保证双方主体数保持一致, 将 m_i 视为与其有同样偏好的 $s(i)$ 个主体, 记为 m_{i_h} , $h=1, 2, \dots, s(i)$, 其中 m_{i_h} 只能匹配一个主体集 W 中的主体。 m_{i_h} 对主体集 W 中各个主体的偏好序值记为: $a_{i_h j} = a_{ij}$, 其中 $h=1, 2, \dots, s(i); j=1, 2, \dots, \max\left(m, \sum_{i=1}^n s(i)\right)$ 。

主体集 W 中各个主体对 m_{i_h} 的偏好序值记为: $b_{i_h j} = b_{ij}$, 其中 $h=1, 2, \dots, s(i); j=1, 2, \dots, \max\left(m, \sum_{i=1}^n s(i)\right)$ 。

这样, 就可以将一对多匹配问题转化为一对一匹配问题进行求解, 下面给出一对多均衡稳定匹配算法。

2. 一对多均衡稳定匹配算法

一对多匹配问题 $\{M, W; \mathbf{P}_M, \mathbf{P}_W, \mathbf{S}\}$ 的均衡稳定匹配的算法可以将其转化为一个等价的一对一匹配问题。构造如下的一对一双边匹配问题 $\{M', W; \mathbf{P}_{M'}, \mathbf{P}_{W'}\}$, 其中:

$M' = \{m'_{1_1}, m'_{1_2}, \dots, m'_{1_{s(1)}}, \dots, m'_{n_1}, m'_{n_2}, \dots, m'_{n_{s(n)}}\}$, m'_{i_h} 与主体集 M 中的 m_i 相对应, $i=1, 2, \dots, n; h=1, 2, \dots, s(i)$;

$\mathbf{P}_{M'} = [a'_{i_h j}]_{k \times k}$, 其中 $a'_{i_h j} = a_{ij}$, $h=1, 2, \dots, s(i); j=1, 2, \dots, \max\left(m, \sum_{i=1}^n s(i)\right); k = \max\left(m, \sum_{i=1}^n s(i)\right)$;

$\mathbf{P}_{W'} = [b'_{i_h j}]_{k \times k}$, 其中 $b'_{i_h j} = b_{ij}$, $h=1, 2, \dots, s(i); j=1, 2, \dots, \max\left(m, \sum_{i=1}^n s(i)\right); k = \max\left(m, \sum_{i=1}^n s(i)\right)$ 。

$\{M, W; \mathbf{P}_M, \mathbf{P}_W, \mathbf{S}\}$ 与 $\{M', W; \mathbf{P}_{M'}, \mathbf{P}_{W'}\}$ 之间的对应关系如下:

设 $\{M, W; \mathbf{P}_M, \mathbf{P}_W, \mathbf{S}\}$ 有一个匹配 $\mu = \{(m_1, w_{1_1}, w_{1_2}, \dots, w_{1_{s(1)}}), \dots, (m_n, w_{n_1}, w_{n_2}, \dots, w_{n_{s(n)}})\}$, 即 $\mu = \{(m_1, w_{1_1}), (m_1, w_{1_2}), \dots, (m_1, w_{1_{s(1)}}), \dots, (m_n, w_{n_1}), (m_n, w_{n_2}), \dots, (m_n, w_{n_{s(n)}})\}$, 其中 $w_{i_h} \in W, w_{i_h}$ 各不相同 ($i=1, 2, \dots, n; h=1, 2, \dots, s(i)$)。与 μ 相对应, 构造 $\{M', W; \mathbf{P}_{M'}, \mathbf{P}_{W'}\}$ 的匹配:

$$\mu' = \{(m'_{1_1}, w_{1_1}), (m'_{1_2}, w_{1_2}), \dots, (m'_{1_{s(1)}}, w_{1_{s(1)}}), \dots, (m'_{n_1}, w_{n_1}), (m'_{n_2}, w_{n_2}), \dots, (m'_{n_{s(n)}}, w_{n_{s(n)}})\}$$

反之亦然。 $\{M, W; \mathbf{P}_M, \mathbf{P}_W, \mathbf{S}\}$ 的一个匹配 μ 是稳定匹配的充分必要条件是 μ 对应的 $\{M', W; \mathbf{P}_{M'}, \mathbf{P}_{W'}\}$ 的一对一匹配 μ' 是稳定匹配^[17]。

一对多双边匹配问题 $\{M, W; P_M, P_W, S\}$ 的均衡稳定匹配算法如下:

步骤 1 构造与 $\{M, W; P_M, P_W, S\}$ 对应的一对一匹配问题 $\{M', W; P_{M'}, P_{W'}\}$ 。

步骤 2 引进 0-1 变量 $x_{i_h j}$ 表示主体集 M' 与 W 的主体对 (m'_{i_h}, w_j) 的状态。若 (m'_{i_h}, w_j) 是匹配对, 则 $x_{i_h j} = 1$, 若 (m'_{i_h}, w_j) 不是匹配对, 则 $x_{i_h j} = 0$, $i = 1, 2, \dots, n; h = 1, 2, \dots, s(i); j = 1, 2, \dots, m$ 。

步骤 3 构建一对一匹配问题 $\{M', W; P_{M'}, P_{W'}\}$ 的均衡匹配模型。

$$\begin{aligned} \min_{x_{i_h j}} \max & \left(\sum_{i=1}^n \sum_{h=1}^{s(i)} \sum_{j=1}^m a'_{i_h j} x_{i_h j}, \sum_{i=1}^n \sum_{h=1}^{s(i)} \sum_{j=1}^m b'_{i_h j} x_{i_h j} \right) \\ \text{s. t.} & \begin{cases} \sum_{j=1}^m x_{i_h j} = 1, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i) \\ \sum_{i=1}^n \sum_{h=1}^{s(i)} x_{i_h j} = 1, & j = 1, 2, \dots, m \\ x_{i_h j} + \sum_{k: a'_{i_h k} \leq a'_{i_h j}} x_{i_h k} + \sum_{l: b'_{l j} \leq b'_{i_h j}} x_{l j} \geq 1, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i); \\ & j = 1, 2, \dots, m \\ x_{i_h j} = \{0, 1\}, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i); \\ & j = 1, 2, \dots, m \end{cases} \end{aligned} \quad (2.2.6)$$

步骤 4 引入变量 U , 将模型(2.2.6)转化为整数线性规划模型:

$$\begin{aligned} \min_{x_{i_h j}} & U \\ \text{s. t.} & \begin{cases} U \geq \sum_{i=1}^n \sum_{h=1}^{s(i)} \sum_{j=1}^m a'_{i_h j} x_{i_h j} \\ U \geq \sum_{i=1}^n \sum_{h=1}^{s(i)} \sum_{j=1}^m b'_{i_h j} x_{i_h j} \\ \sum_{j=1}^m x_{i_h j} = 1, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i) \\ \sum_{i=1}^n \sum_{h=1}^{s(i)} x_{i_h j} = 1, & j = 1, 2, \dots, m \\ x_{i_h j} + \sum_{k: a'_{i_h k} \leq a'_{i_h j}} x_{i_h k} + \sum_{l: b'_{l j} \leq b'_{i_h j}} x_{l j} \geq 1, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i); \\ & j = 1, 2, \dots, m \\ x_{i_h j} = \{0, 1\}, & i = 1, 2, \dots, n; h = 1, 2, \dots, s(i); \\ & j = 1, 2, \dots, m \end{cases} \end{aligned} \quad (2.2.7)$$

步骤 5 求解模型(2.2.7)得最优解 $x_{i_h j}^*$, $i=1, 2, \dots, n; h=1, 2, \dots, s(i); j=1, 2, \dots, m$;

步骤 6 写出最优解 $x_{i_h j}^*$ 对应 $\{M', W; P_{M'}, P_{W'}\}$ 的均衡匹配:

$$\mu'^* = \{(m'_{1_1}, \omega_{1_1}^*), (m'_{1_2}, \omega_{1_2}^*), \dots, (m'_{1_{s(1)}}, \omega_{1_{s(1)}}^*), \dots, (m'_{n_1}, \omega_{n_1}^*), \\ (m'_{n_2}, \omega_{n_2}^*), \dots, (m'_{n_{s(n)}}, \omega_{n_{s(n)}}^*)\}$$

步骤 7 根据 $\{M, W; P_M, P_W, S\}$ 与 $\{M', W; P_{M'}, P_{W'}\}$ 的对应关系, 将 μ'^* 转化为 $\{M, W; P_M, P_W, S\}$ 的均衡匹配:

$$\mu^* = \{(m_1, \omega_{1_1}^*), (m_1, \omega_{1_2}^*), \dots, (m_1, \omega_{1_{s(1)}}^*), \dots, (m_n, \omega_{n_1}^*), \\ (m_n, \omega_{n_2}^*), \dots, (m_n, \omega_{n_{s(n)}}^*)\}$$

算法结束。

专家和患者是预约的双方。设某医院在某时间段内有 m 位患者 $W = \{\omega_1, \omega_2, \dots, \omega_m\}$ 挂号专家门诊。先由患者填写问卷量表, 该量表在疾病知识库的支持下, 比对患者的主诉、现病史、既往史、家族史、婚育情况、疫区疫水接触史等信息, 推断患者疾病所属的专科科室; 系统推出该专科科室内 n 位专家 $M = \{m_1, m_2, \dots, m_n\}$ 及其诊疗特长的简介, 包括专家 m_i 在这时间段内可看诊人数 $s(i)$, $i=1, 2, \dots, n$, 由此患者提出选择专家的意向矩阵 P_W 。同时, 系统把该专科内专家诊疗的特长与患者进行比对, 得到专家对患者的意向矩阵 P_M 。把专家和患者看成是一对多匹配问题 $\{M, W; P_M, P_W, S\}$ 的主体集 M 和 W , 由一对多匹配算法计算, 得到患者和专家预约的结果, 供患者和专家双方认可。整个预约过程如图 2-2 所示。

3. 专家门诊的匹配预约

系统输出得到匹配的专家和患者(即预约成功的专家和患者), 并更新原系统的专家集和患者集, 如果新的专家集或者新的患者集中有一个是空集, 则算法终止, 否则把新的专家集和新的患者集输入系统, 进行再次匹配预约。一对多预约匹配算法如图 2-3 所示。

2.2.4 算例

设某医院某专科有 $n=4$ 位专家, $M = \{m_1, m_2, m_3, m_4\}$, 每位专家在某时间段可以诊疗 4 位患者($s(i)=4, i=1, 2, 3, 4$), 在该时间段内挂该科专家门诊的患者数 $m=14$ 人, $W = \{\omega_1, \omega_2, \dots, \omega_{14}\}$, 专家 M 对患者 W 的意向矩阵 P_M 为

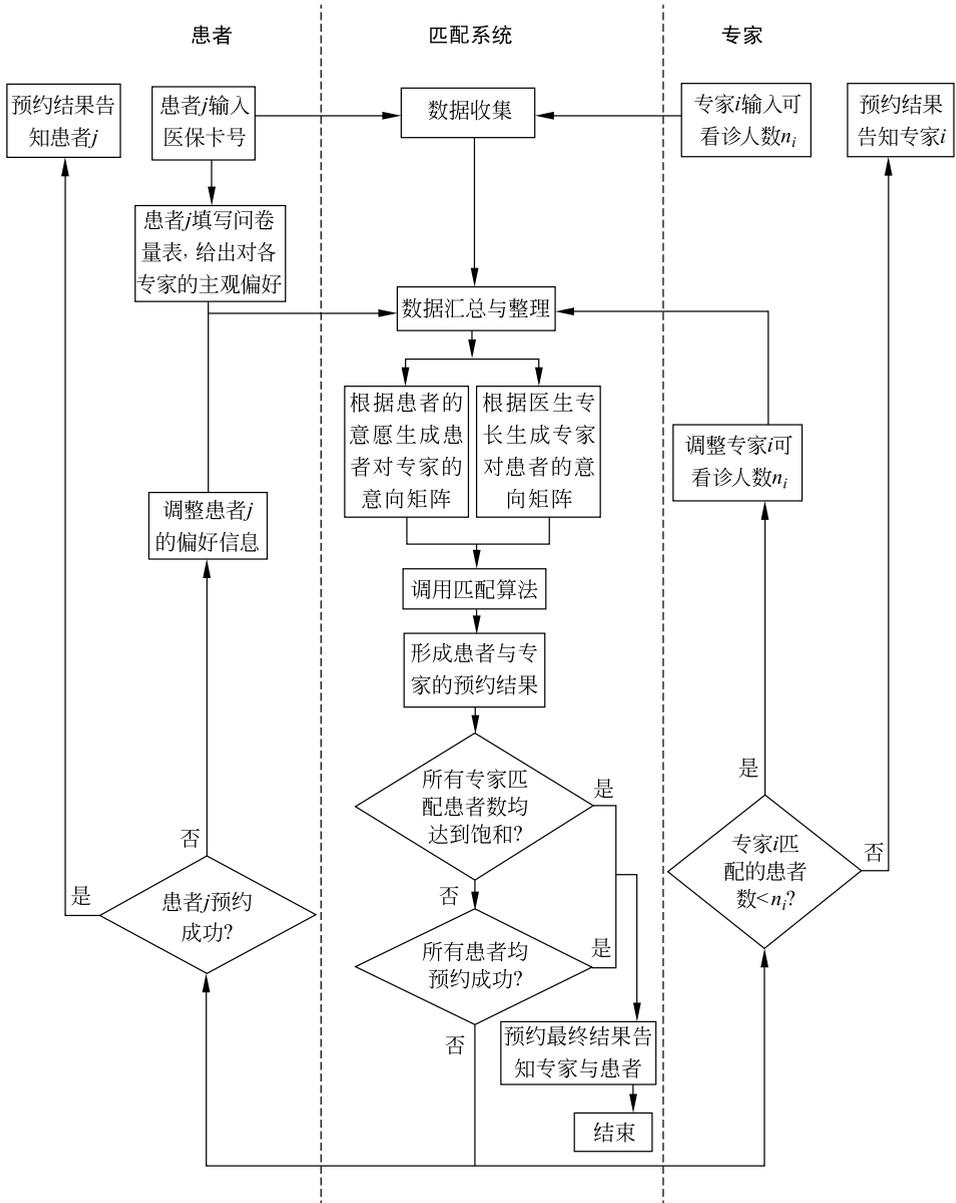


图 2-2 预约流程图

$$P_M = [a_{ij}]_{4 \times 14} = \begin{vmatrix} 1 & 2 & 1 & 1 & 3 & 2 & 2 & 2 & 2 & 3 & 2 & 3 & 3 & 3 \\ 2 & 3 & 3 & 3 & 4 & 1 & 4 & 2 & 1 & 2 & 2 & 3 & 2 & 3 \\ 1 & 2 & 1 & 3 & 2 & 2 & 3 & 3 & 2 & 3 & 3 & 1 & 3 & 1 \\ 4 & 2 & 4 & 3 & 1 & 2 & 2 & 4 & 4 & 4 & 3 & 3 & 1 & 1 \end{vmatrix}$$

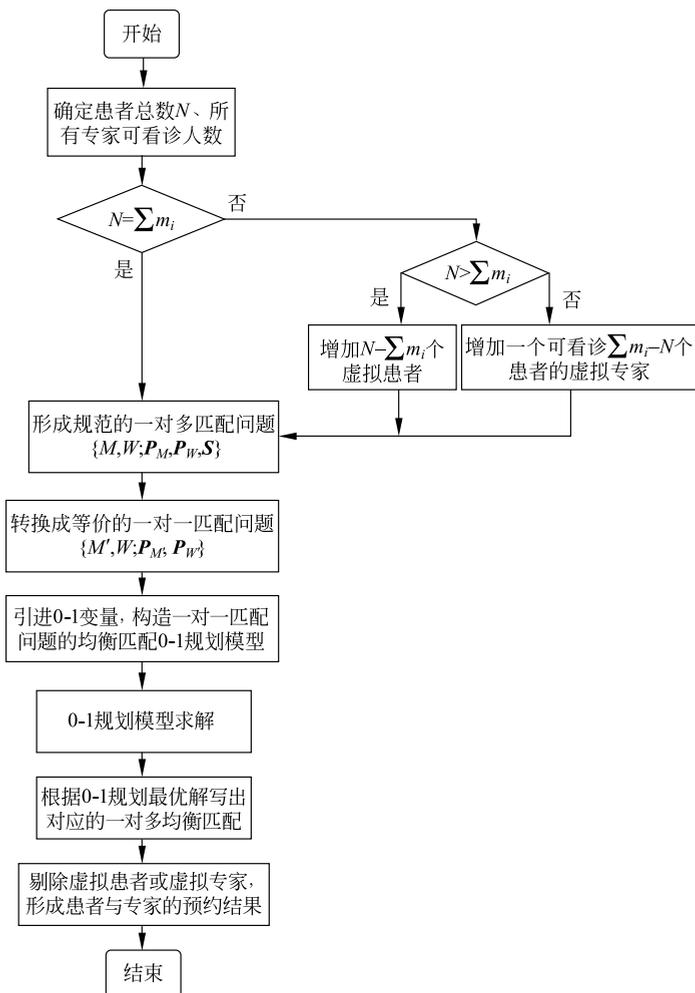


图 2-3 一对多匹配算法

其中, a_{ij} 表示专家 m_i 合适诊疗患者 w_j 的意向序值。如 $a_{11} = a_{13} = a_{14} = 1$ 表示对专家 m_1 而言, 患者 w_1, w_3, w_4 的病症类型相似, 都是专家 m_1 第 1 意愿诊疗的病症, $a_{12} = a_{16} = a_{17} = a_{18} = a_{19} = a_{1,11} = 2$ 表示患者 $w_2, w_6, w_7, w_8, w_9, w_{11}$ 的病症同类, 是专家 m_1 第 2 意愿诊疗的病症, 等等。

患者 W 对专家 M 的意向矩阵 P_W 为

$$P_W = [b_{ij}]_{4 \times 14} = \begin{vmatrix} 1 & 3 & 4 & 2 & 4 & 4 & 3 & 3 & 3 & 2 & 4 & 3 & 4 & 2 \\ 3 & 2 & 3 & 3 & 3 & 2 & 4 & 2 & 4 & 4 & 1 & 4 & 2 & 1 \\ 4 & 4 & 1 & 4 & 2 & 1 & 2 & 4 & 2 & 3 & 2 & 1 & 3 & 3 \\ 2 & 1 & 2 & 1 & 1 & 3 & 1 & 1 & 1 & 1 & 3 & 2 & 1 & 4 \end{vmatrix}$$

其中, b_{ij} 表示患者 w_j 希望专家 m_i 给其诊疗的意向序值。如该矩阵的第一列数字(1,3,4,2)表示患者 w_1 最希望 1 号专家给他诊疗,其次是 3 号专家,再次是 4 号专家,最后是 2 号专家。

由于患者总数 $m=14 < \sum_{i=1}^4 s_i=16$, 为此引进 2 个虚拟患者 w_{15} 和 w_{16} , 并认为专家对这些虚拟患者诊疗的意向序值是 $a_{ij}=15, i=1,2,3,4; j=15,16$, 而这些虚拟患者希望各个专家诊疗的意向序值都是 1, 即 $b_{ij}=1, i=1,2,3,4; j=15,16$ 。

现对矩阵 \mathbf{P}_M 和 \mathbf{P}_W 进行一致性处理, 在保证原有矩阵结构不变的基础上使 \mathbf{P}_M 的每一行之和都为 $1+2+\dots+m=1+2+\dots+16=136$, \mathbf{P}_W 的每一列序值之和都为 $1+2+\dots+n=1+2+3+4=10$, 得到一致性处理后的意向矩阵。

$\mathbf{P}_M =$

$$\begin{bmatrix} 2 & 6.5 & 2 & 2 & 12 & 6.5 & 6.5 & 6.5 & 6.5 & 12 & 6.5 & 12 & 12 & 12 & 15.5 & 15.5 \\ 5 & 10 & 10 & 10 & 13.5 & 1.5 & 13.5 & 5 & 1.5 & 5 & 5 & 10 & 5 & 10 & 15.5 & 15.5 \\ 2.5 & 6.5 & 2.5 & 11.5 & 6.5 & 6.5 & 11.5 & 11.5 & 6.5 & 11.5 & 11.5 & 2.5 & 11.5 & 2.5 & 15.5 & 15.5 \\ 12 & 5 & 12 & 8 & 2 & 5 & 5 & 12 & 12 & 12 & 8 & 8 & 2 & 2 & 15.5 & 15.5 \end{bmatrix}$$

$$\mathbf{P}_W = \begin{bmatrix} 1 & 3 & 4 & 2 & 4 & 4 & 3 & 3 & 3 & 2 & 4 & 3 & 4 & 2 & 2.5 & 2.5 \\ 3 & 2 & 3 & 3 & 3 & 2 & 4 & 2 & 4 & 4 & 1 & 4 & 2 & 1 & 2.5 & 2.5 \\ 4 & 4 & 1 & 4 & 2 & 1 & 2 & 4 & 2 & 3 & 2 & 1 & 3 & 3 & 2.5 & 2.5 \\ 2 & 1 & 2 & 1 & 1 & 3 & 1 & 1 & 1 & 1 & 3 & 2 & 1 & 4 & 2.5 & 2.5 \end{bmatrix}$$

将以上一对多匹配问题转化为一对一匹配问题, 其中:

$\mathbf{P}_{M'} = [a_{ij}]_{16 \times 16} =$

$$\begin{bmatrix} 2 & 6.5 & 2 & 2 & 12 & 6.5 & 6.5 & 6.5 & 6.5 & 12 & 6.5 & 12 & 12 & 12 & 15.5 & 15.5 \\ 2 & 6.5 & 2 & 2 & 12 & 6.5 & 6.5 & 6.5 & 6.5 & 12 & 6.5 & 12 & 12 & 12 & 15.5 & 15.5 \\ 2 & 6.5 & 2 & 2 & 12 & 6.5 & 6.5 & 6.5 & 6.5 & 12 & 6.5 & 12 & 12 & 12 & 15.5 & 15.5 \\ 2 & 6.5 & 2 & 2 & 12 & 6.5 & 6.5 & 6.5 & 6.5 & 12 & 6.5 & 12 & 12 & 12 & 15.5 & 15.5 \\ 5 & 10 & 10 & 10 & 13.5 & 1.5 & 13.5 & 5 & 1.5 & 5 & 5 & 10 & 5 & 10 & 15.5 & 15.5 \\ 5 & 10 & 10 & 10 & 13.5 & 1.5 & 13.5 & 5 & 1.5 & 5 & 5 & 10 & 5 & 10 & 15.5 & 15.5 \\ 5 & 10 & 10 & 10 & 13.5 & 1.5 & 13.5 & 5 & 1.5 & 5 & 5 & 10 & 5 & 10 & 15.5 & 15.5 \\ 5 & 10 & 10 & 10 & 13.5 & 1.5 & 13.5 & 5 & 1.5 & 5 & 5 & 10 & 5 & 10 & 15.5 & 15.5 \\ 2.5 & 6.5 & 2.5 & 11.5 & 6.5 & 6.5 & 11.5 & 11.5 & 6.5 & 11.5 & 11.5 & 2.5 & 11.5 & 2.5 & 15.5 & 15.5 \\ 2.5 & 6.5 & 2.5 & 11.5 & 6.5 & 6.5 & 11.5 & 11.5 & 6.5 & 11.5 & 11.5 & 2.5 & 11.5 & 2.5 & 15.5 & 15.5 \\ 2.5 & 6.5 & 2.5 & 11.5 & 6.5 & 6.5 & 11.5 & 11.5 & 6.5 & 11.5 & 11.5 & 2.5 & 11.5 & 2.5 & 15.5 & 15.5 \\ 2.5 & 6.5 & 2.5 & 11.5 & 6.5 & 6.5 & 11.5 & 11.5 & 6.5 & 11.5 & 11.5 & 2.5 & 11.5 & 2.5 & 15.5 & 15.5 \\ 12 & 5 & 12 & 8 & 2 & 5 & 5 & 12 & 12 & 12 & 8 & 8 & 2 & 2 & 15.5 & 15.5 \\ 12 & 5 & 12 & 8 & 2 & 5 & 5 & 12 & 12 & 12 & 8 & 8 & 2 & 2 & 15.5 & 15.5 \\ 12 & 5 & 12 & 8 & 2 & 5 & 5 & 12 & 12 & 12 & 8 & 8 & 2 & 2 & 15.5 & 15.5 \\ 12 & 5 & 12 & 8 & 2 & 5 & 5 & 12 & 12 & 12 & 8 & 8 & 2 & 2 & 15.5 & 15.5 \end{bmatrix}$$

$$P_{W'} = [b_{ij}]_{16 \times 16} =$$

1	3	4	2	4	4	3	3	3	2	4	3	4	2	2.5	2.5
1	3	4	2	4	4	3	3	3	2	4	3	4	2	2.5	2.5
1	3	4	2	4	4	3	3	3	2	4	3	4	2	2.5	2.5
1	3	4	2	4	4	3	3	3	2	4	3	4	2	2.5	2.5
3	2	3	3	3	2	4	2	4	4	1	4	2	1	2.5	2.5
3	2	3	3	3	2	4	2	4	4	1	4	2	1	2.5	2.5
3	2	3	3	3	2	4	2	4	4	1	4	2	1	2.5	2.5
3	2	3	3	3	2	4	2	4	4	1	4	2	1	2.5	2.5
4	4	1	4	2	1	2	4	2	3	2	1	3	3	2.5	2.5
4	4	1	4	2	1	2	4	2	3	2	1	3	3	2.5	2.5
4	4	1	4	2	1	2	4	2	3	2	1	3	3	2.5	2.5
4	4	1	4	2	1	2	4	2	3	2	1	3	3	2.5	2.5
2	1	2	1	1	3	1	1	1	1	3	2	1	4	2.5	2.5
2	1	2	1	1	3	1	1	1	1	3	2	1	4	2.5	2.5
2	1	2	1	1	3	1	1	1	1	3	2	1	4	2.5	2.5
2	1	2	1	1	3	1	1	1	1	3	2	1	4	2.5	2.5

对于上述一对多双边匹配问题 $\{M, W, P_M, P_W, S\}$, 我们用 4 个算法进行计算和比较。

1. 有利于专家的稳定匹配

以专家满意度总和最小作为目标函数, 建立如下模型:

$$\min_{x_{ij}} \left\{ \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij} \right\}$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\ x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{cases}$$

(2.2.8)

这是 0-1 线性规划。

采用 Lingo16 软件求解以上模型,得到稳定匹配是:

$$\mu'_1 = \{(m'_1, \omega_8), (m'_2, \omega_3), (m'_3, \omega_1), (m'_4, \omega_4), (m'_5, \omega_9), (m'_6, \omega_6), \\ (m'_7, \omega_{11}), (m'_8, \omega_{10}), (m'_9, \omega_{12}), (m'_{10}, \omega_{16}), (m'_{11}, \omega_{15}), (m'_{12}, \omega_{14}), \\ (m'_{13}, \omega_{13}), (m'_{14}, \omega_{15}), (m'_{15}, \omega_2), (m'_{16}, \omega_7)\}$$

以上 m'_1, m'_2, m'_3, m'_4 对应专家 m_1 ; m'_5, m'_6, m'_7, m'_8 对应专家 m_2 ; $m'_9, m'_{10}, m'_{11}, m'_{12}$ 对应专家 m_3 ; $m'_{13}, m'_{14}, m'_{15}, m'_{16}$ 对应专家 m_4 ; ω_{15}, ω_{16} 为虚拟患者。所以有利于专家的稳定匹配为

$$\mu_1 = \{(m_1, \omega_8), (m_1, \omega_3), (m_1, \omega_1), (m_1, \omega_4), (m_2, \omega_9), (m_2, \omega_6), \\ (m_2, \omega_{11}), (m_2, \omega_{10}), (m_3, \omega_{12}), (m_3, \omega_{14}), (m_4, \omega_{13}), (m_4, \omega_2), \\ (m_4, \omega_7)\}$$

在匹配 μ_1 中,专家的意向序值总和是 $\sum_{(m_i, \omega_j) \in \mu_1} a_{ij} = 44.5$, 患者的意向序值总和是 $\sum_{(m_i, \omega_j) \in \mu_1} b_{ij} = 29$, 两者总和 $\sum_{(m_i, \omega_j) \in \mu_1} a_{ij} + \sum_{(m_i, \omega_j) \in \mu_1} b_{ij} = 44.5 + 29 = 73.5$, 两者之差 $\left| \sum_{(m_i, \omega_j) \in \mu_1} a_{ij} - \sum_{(m_i, \omega_j) \in \mu_1} b_{ij} \right| = |44.5 - 29| = 15.5$ 。

2. 有利于患者的稳定匹配

以患者满意度总和最小作为目标函数,建立如下模型:

$$\min \left\{ \sum_{i,j}^n \sum_{i,j}^n b_{ij} x_{ij} \right\}$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\ x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{cases}$$

(2.2.9)

这是 0-1 线性规划。

采用 Lingo16 软件求解以上模型,得到稳定匹配是:

$$\mu'_2 = \{(m'_1, \omega_1), (m'_2, \omega_{16}), (m'_3, \omega_{10}), (m'_4, \omega_4), (m'_5, \omega_8), (m'_6, \omega_{11}), \\ (m'_7, \omega_{15}), (m'_8, \omega_{14}), (m'_9, \omega_3), (m'_{10}, \omega_{12}), (m'_{11}, \omega_5), (m'_{12}, \omega_6), \\ (m'_{13}, \omega_2), (m'_{14}, \omega_{13}), (m'_{15}, \omega_9), (m'_{16}, \omega_7)\}$$

同上,经过整理后得到有利于患者的稳定匹配为

$$\mu_2 = \{(m_1, \tau_1), (m_1, \tau_{10}), (m_1, \tau_4), (m_2, \tau_8), (m_2, \tau_{11}), (m_2, \tau_{14}), \\ (m_3, \tau_3), (m_3, \tau_{12}), (m_3, \tau_5), (m_3, \tau_6), (m_4, \tau_2), (m_4, \tau_{13}), \\ (m_4, \tau_9), (m_4, \tau_7)\}$$

在匹配 μ_2 中,专家的意向序值总和是 $\sum_{(m_i, \tau_j) \in \mu_2} a_{ij} = 83$, 患者的意向序值总和是 $\sum_{(m_i, \tau_j) \in \mu_2} b_{ij} = 18$, 两者之和是 $\sum_{(m_i, \tau_j) \in \mu_2} a_{ij} + \sum_{(m_i, \tau_j) \in \mu_2} b_{ij} = 83 + 18 = 101$; 两者之差是 $\left| \sum_{(m_i, \tau_j) \in \mu_2} a_{ij} - \sum_{(m_i, \tau_j) \in \mu_2} b_{ij} \right| = |83 - 18| = 65$ 。

3. 考虑双方满意度之和最小化

以专家满意度和患者满意度总和最小作为目标函数,建立如下模型:

$$\min_{x_{ij}} \left\{ \sum_{i=1}^n \sum_{j=1}^n (a_{ij} + b_{ij}) x_{ij} \right\}$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} + \sum_{k: a_{ik} \leq a_{ij}} x_{ik} + \sum_{l: b_{lj} \leq b_{ij}} x_{lj} \geq 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \\ x_{ij} = \{0, 1\}, & i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{cases}$$

(2.2.10)

采用 Lingo16 软件求解以上模型,得到稳定匹配是:

$$\mu'_3 = \{(m'_1, \tau_1), (m'_2, \tau_{16}), (m'_3, \tau_8), (m'_4, \tau_4), (m'_5, \tau_{11}), (m'_6, \tau_9), \\ (m'_7, \tau_6), (m'_8, \tau_{10}), (m'_9, \tau_3), (m'_{10}, \tau_{12}), (m'_{11}, \tau_{15}), (m'_{12}, \tau_{14}), \\ (m'_{13}, \tau_{13}), (m'_{14}, \tau_5), (m'_{15}, \tau_2), (m'_{16}, \tau_7)\}$$

经过整理后得到对应稳定匹配为

$$\mu_3 = \{(m_1, \tau_1), (m_1, \tau_8), (m_1, \tau_4), (m_2, \tau_{11}), (m_2, \tau_9), (m_2, \tau_6), \\ (m_2, \tau_{10}), (m_3, \tau_3), (m_3, \tau_{12}), (m_3, \tau_{14}), (m_4, \tau_{13}), (m_4, \tau_5), \\ (m_4, \tau_2), (m_4, \tau_7)\}$$

在匹配 μ_3 中,专家的意向序值总和是 $\sum_{(m_i, \tau_j) \in \mu_3} a_{ij} = 45$, 患者的意向序值

总和是 $\sum_{(m_i, w_j) \in \mu_3} b_{ij} = 26$, 两者之和是 $\sum_{(m_i, w_j) \in \mu_3} a_{ij} + \sum_{(m_i, w_j) \in \mu_3} b_{ij} = 45 + 26 = 71$; 两者之差是 $\left| \sum_{(m_i, w_j) \in \mu_3} a_{ij} - \sum_{(m_i, w_j) \in \mu_3} b_{ij} \right| = |45 - 26| = 19$ 。

4. 均衡稳定匹配

由式(2.2.7)可以求得本问题的均衡稳定匹配是:

$$\mu^* = \{(m_1, w_{11}), (m_1, w_3), (m_1, w_1), (m_1, w_4), (m_2, w_8), (m_2, w_6), (m_2, w_9), (m_2, w_{10}), (m_3, w_{12}), (m_3, w_{14}), (m_4, w_{13}), (m_4, w_7), (m_4, w_5), (m_4, w_2)\}$$

这时 4 位专家的意向序值总和为 $\sum_{(m_i, w_j) \in \mu^*} a_{ij} = 44.5$, 14 位患者的意向序值总和为 $\sum_{(m_i, w_j) \in \mu^*} b_{ij} = 41$, 两者之和 $\sum_{(m_i, w_j) \in \mu^*} a_{ij} + \sum_{(m_i, w_j) \in \mu^*} b_{ij} = 44.5 + 41 = 85.5$, 两者之差 $\left| \sum_{(m_i, w_j) \in \mu^*} a_{ij} - \sum_{(m_i, w_j) \in \mu^*} b_{ij} \right| = |44.5 - 41| = 3.5$ 。

以上 4 种算法得到的结果总结如表 2-1 所示。

表 2-1 4 种算法结果对比

参数	$\sum_{(m_i, w_j) \in \mu^*} a_{ij}$	$\sum_{(m_i, w_j) \in \mu^*} b_{ij}$	$\sum_{(m_i, w_j) \in \mu^*} a_{ij} + \sum_{(m_i, w_j) \in \mu^*} b_{ij}$	$\left \sum_{(m_i, w_j) \in \mu^*} a_{ij} - \sum_{(m_i, w_j) \in \mu^*} b_{ij} \right $
有利于专家	44.5	29	73.5	15.5
有利于患者	83	18	101	65
考虑双方满意度之和最小	45	26	71	19
均衡稳定匹配	44.5	41	85.5	3.5

上述算例表明,均衡稳定算法得到的匹配 μ^* ,其专家意向序值为 44.5,是 4 种方法中最小的;匹配 μ^* 的专家与患者总意向之差为 3.5,是 4 种方法中最小的。因此说,均衡匹配能够较好地考虑专家和患者双方的偏好。

采取患者优先选择专家的门诊预约会降低专家在门诊预约中的主导性,即:专家所看的患者可能不是他所偏好的,也不是他专长的。专家优先选择患者的门诊预约也会导致患者的选择权得不到充分的保障,给患者看病的专家可

能并不是该患者中意的。由此可见,均衡匹配算法得到的匹配结果较好地平衡了专家特长的诊疗有效性和患者选择专家意向性之间的矛盾。

2.2.5 管理启示

考虑专家特长和患者偏好的匹配预约不仅改变了以患者单方意愿择医,也改变了以往门诊预约开放的时间。由于匹配预约的结果需要得到患者和专家的认可,必须留有适当的时间,所以预约将在门诊时间前7天开始直至门诊时间前1天截止。这些改变对于居民的就医行为、满意度和爽约率带来怎样的影响,是需要进一步研究的;其次,匹配模型推荐的专家基本上是适合患者病情的,然而患者择医往往受到专家的声誉、职称或心理因素的影响,匹配预约提出的专家可能因患者不认可而产生算法“死循环”。怎样避免“死循环”,怎么跳出“死循环”也是需要进一步研究和分析的;最后,匹配模型的前提条件是供小于求,即患者第一顺位选择的专家资源是稀缺的,医院只优先满足病情匹配的患者的需求,但如果专家资源富裕并供大于求,算法很快就终止。

匹配预约的成效可以通过双方的满意度来评价。患者方的满意度包括两方面:一是其主观意向的满意度,即患者根据主观意愿选择专家的成功率,用预约成功率可以评价;二是其病情的客观诊疗需求与专家的擅长疾病类型匹配程度,可以通过预约成功后的病种-专家特长符合率来评价。专家方面的满意度是考量专家是否收治了符合自身专业方向与疾病特长的患者,可以通过门诊目标病种收治率来评价,同时可用门诊住院比作为参考指标。

中国人口总量增加、人口老龄化和患病率上升等因素使得大型医院的门诊压力不断增大。预约作为有效利用资源和提高服务质量的手段受到越来越多的重视。医院实施门诊预约,患者在挂号时,由于对自身病情判断不准和对专家特长不清楚,盲目地倾向于选择名家,从而使得专家门诊预约分布不均衡,导致专家资源未能合理利用。为了让患者找到合适的专家,也让专家看诊自己擅长的病症,本部分基于双边匹配理论提出了专家门诊匹配预约机制,建立专家-患者的一对多匹配预约模型。为了体现匹配的“公平性”,让专家、患者双方都能表达自己的要求,提出专家-患者均衡匹配。算例表明,如果患者优先选专家,那么预约会降低专家的指导性;如果专家优先选患者,那么预约会导致患者丧失选择权。均衡匹配算法得到的匹配结果能较好地平衡专家特长与患者选择专家的意向性,提高双方的满意度。

本部分提出均衡匹配预约在实际中应用还需要:

(1)设计患者预检的量表,搜集患者病症的基本特征,以判断什么样的专家更适合该患者,搜集患者对专家的意向和选择;

(2)建立专家库,记录每个专家的特长。

2.3 考虑患者行为的门诊资源配置

2.3.1 相关研究概述

在中国,“看病难,看病贵”成为困扰普通老百姓的一个大问题。“看病难”是由于医疗资源配置严重失衡,优质医疗资源过度集中在二、三级医院所导致的^[19]。优秀的医生和优良的设备集中在二级、三级医院,导致“大医院爆满,小医院冷清”^[20]。根据国家卫生健康委员会统计信息中心统计数据,2022年,全国医疗卫生机构总诊疗人次达84.16亿,其中三级医院诊疗人次约占26.48%;从看病费用方面看,2022年,全国三级公立医院次均门诊费用为381.6元,二级公立医院次均门诊费用为241.2元(以上数据均来源于国家卫生健康委员会统计信息中心网站)。另外根据上海市卫生健康委员会发布最新统计数据显示:2022年度,本市门急诊人次达21959.68万人次。其中综合医院门急诊量占全市门急诊总量的49.17%,社区卫生服务中心门急诊量占比25.25%。2014年医药产业调查表明,全国三级甲等医院每天接诊的患者中,近70%是可以到社区医院或者一、二级医院接受治疗的^[21]。根据2024年08月21日《健康报》报道,2023年,全国基层医疗卫生机构门急诊诊疗人次占比达52%,医疗资源纵向流动、患者双向转诊成为新常态。我国分级诊疗体系建设虽然已取得积极进展和明显成效,但是与人民群众期盼还有距离。

既然三级医院排队长,看病贵,为何人们依然愿意选择去大医院看病?张春瑜等^[22]在某三级甲等医院的患者中选择1443人进行问卷调查,调查研究发现,影响患者选择的主要因素分3类:患者认知、相关群体和媒体因素。在认知方面,患者选择该医院就医的考虑因素依次是医疗技术水平、位置远近、服务态度、有无熟人、就医环境、价格。其中,以医疗技术水平作为首要考虑因素的患者最多,占69.50%;将医疗价格作为首要考虑因素的患者仅占0.42%;其他将距离近、服务态度好、有熟人、就医环境好作为首要考虑因素的患者分别占16.63%、7.90%、3.47%、2.08%。由此可见,大部分患者在选择时首要考虑的还是医疗技术水平,而专家门诊则普遍被认为是“信得过”的。所以,大量患者涌向三级甲等医院预约专家门诊,在这个过程中,患者不仅受到医疗技术水平、就医环境等因素的影响,还与其对自身病情的判断和对不同医疗机构的认知、期望有关。

国外在20世纪90年代就有学者开始研究患者选择医院的问题,比较典型的文献有Adams等^[23]研究发现农村市场医疗保险受益人在医院选择时,距离

不是最重要的因素;Bronstein等^[24]研究发现产妇在选择产科护理时不会选择农村医院。近年的研究有:Ali^[25]采用调研的方法研究了患者在选择医院时最重要的几个因素;Smith等^[26]使用多项式逻辑回归方法对患者选择医院问题建立模型;Walter^[27]设计了一个两阶段模型,包含患者和全科医生选择优化。另外还有一些文献研究医疗保险、公立/私立医院等因素选择医院的问题。

综合以上,目前已有大量文献针对患者在选择医院时的影响因素进行研究,研究方法大多采用调研方法,结合统计学方法建立模型,从微观角度研究患者在选择医院时的行为决策的文献非常少。

针对目前患者盲目选择“最好的医院”行为,各级医疗机构都在推行分级诊疗机制。在此过程中,最重要的就是对患者进行合理分流。如果要对患者进行分流,则需要充分了解患者在医院选择时考虑的因素和行为决策特征,针对其考虑因素和行为决策特征进行合理引导。而目前患者选择医院时一个重要体现就是专家门诊的选择与预约。本节对专家门诊预约中患者行为决策模式进行分析,建立行为决策模型,从而帮助医院建立合适的应对策略,合理配置门诊资源,为进一步提出患者分流策略提供参考。

2.3.2 问题描述与分析

1. 预约排队行为分析

选择医疗机构的人员分为两类:一是患者自己;二是患者家属。具有行为能力、非重病的患者一般能够自主做出决策,而不具有行为能力的,如孩子和老人或是重病患者,一般无法自己做出决策,而是由家属做出决策。以下分别对其预约行为进行分析:

一是患者自己做决策的情形。这一类患者对自己的病情有较清晰的认知,能根据获得的信息做出相对客观和理性的判断。在对医疗服务机构和医生的选择上,更趋于理性判断,根据病情轻重、医疗服务获得的难易程度等综合考虑进行选择。

二是患者家属做决策。这种情况比较复杂,受到患者家属焦虑程度影响很大。患者本身无法非常清楚地表达自己的病情情况,即使能够表达,家属无法亲身感受其痛苦但内心非常焦虑,导致对病情过度估计的情形非常多。在这种情绪影响下,在选择医疗服务机构时带有很盲目的成分,一味地追求最好的医生、最好的医院。最典型的例子是三级甲等医院的儿科,总是人满为患。当小孩感冒发烧时,家长很焦急,频繁地到医院,导致儿科排长队。

由以上两种行为分析发现,影响患者预约专家门诊的重要因素是其对病情的估计和对医疗服务的期望质量判断。

2. 专家的服务质量分析

患者的病情估计决定预约时的焦虑程度,焦虑程度越高,则对专家门诊服务质量期望越高。所以,病情估计可以用焦虑程度来衡量,焦虑程度和对专家门诊服务期望质量是正相关的。

医院里的专家是有特殊知识、技能和责任的,其服务质量可以由他所具有的知识和技能来体现。而患者得到的服务质量可以由专家的等级和服务时间两个因素来度量^[28]。在此基础上,专家的口碑或声誉也是影响其服务质量的一个因素。Bansal 等^[29]通过实验证明:顾客感知的企业服务质量越高,顾客越有可能产生正面的口碑传播,向他人推荐服务。因而,服务质量与口碑传播之间具有正效应。黄孝俊等^[30]研究结果表明:正面的口碑传播在影响消费者购买决策的同时,也提高了消费者对该产品及其服务的预期。李林等^[31]研究结果表明:顾客感知质量是良好的,倾向于传播正面的口碑,反之,感知质量是低下的,顾客倾向于传播负面的口碑。由以上研究可见,患者所感知的专家服务质量受到专家口碑或声誉的影响,所以应该将这个因素纳入服务质量模型。

对 Anand^[32]和周华^[28]的服务质量模型进行进一步修改,建立如下服务质量模型,以表示服务质量、专家等级、服务时间和专家口碑之间的关系:

$$Q(z) = [Q_b - z\mu]^+ \alpha = \left[Q_b - \frac{z}{t_1} \right]^+ \alpha \quad (2.3.1)$$

式中, $[Q_b - z\mu]^+ = \max(Q_b - z\mu, 0)$, z 表示专家的等级,数字越小,等级越高; $Q(z)$ 表示该专家的服务质量; Q_b 表示患者所能获得的最大服务质量,由最大服务率、最高专家等级、最好口碑构成; μ 表示该专家的平均服务率, $\mu = 1/t_1$, t_1 为该专家的平均服务时间,当 $t_1 \rightarrow \infty$ 时,对应的 $\frac{z}{t_1} \rightarrow 0$, 则有 $Q(z) \rightarrow Q_b$; α 代表专家的口碑, $\alpha \in [0, 1]$ 。从式(2.3.1)中可知:专家的服务质量随服务时间的增加而提高,但提高的幅度呈递减趋势,因为服务质量是有上限的,不能无限提高。而当服务时间相等时,等级高的专家能够提供更高的服务质量;当服务时间越来越长时,不同等级专家所提供的服务质量的差距越来越小。 α 越大,服务质量越高,当专家的口碑非常差,达到 0 时,将使得服务质量为 0,没有患者愿意预约该专家。

3. 患者排队的成本分析

患者的排队成本由 3 个部分构成:等待成本、专家门诊价格和沉没成本。

等待成本是等待时间 t_2 的线性函数,用 $c(t_2)$ 表示。假设患者的到来服从泊松分布,到来率用 λ 来表示;而专家的服务率用 μ 来表示, $\mu = \frac{1}{t_1}$ 。从而等待

时间 t_2 的期望值为

$$E(t_2) = \begin{cases} \frac{1}{\mu - \lambda} = \frac{t_1}{1 - \lambda t_1}, & 0 \leq \lambda t_1 \leq 1 \\ \infty, & \lambda t_1 > 1 \end{cases} \quad (2.3.2)$$

期望等待成本为

$$E(c(t_2)) = \begin{cases} \frac{c}{\mu - \lambda} = \frac{ct_1}{1 - \lambda t_1}, & 0 \leq \lambda t_1 \leq 1 \\ \infty, & \lambda t_1 > 1 \end{cases} \quad (2.3.3)$$

式(2.3.3)中的 c 表示单位等待成本。当到达率超过服务率时,队伍将不断延长,所以等待时间和等待成本趋向于无穷大。

专家门诊价格与专家等级相关,是专家等级的线性函数,用 $p(z)$ 表示, $p(z) = p(z_{\max} + 1 - z)$,其中 p 为对应等级专家单位服务价格, z_{\max} 为专家等级最大取值,取值越大,等级越低。

沉没成本是指由于过去的决策已经发生,而不能由现在或将来的任何决策改变的成本。患者在选择某个专家门诊预约排队等待前已经花费了一定时间成本、金钱成本或体力成本,这部分计入沉没成本。用 c_s 来表示。等待时间越长,沉没成本越高,设 $c_s = bt_2$,其中 b 为单位等待时间成本。

考虑每个患者对不同成本因素的敏感程度不同,设置 $w_i \in [0, 1], i = 1, 2, 3$ 作为每个成本因素的权重,其中 $\sum_{i=1}^3 w_i = 1$ 。

所以排队成本可以表示为

$$T_c = w_1 c(t_2) + w_2 p(z) + w_3 c_s \quad (2.3.4)$$

4. 焦虑程度分析

焦虑程度用 $A(r, e)$ 表示,主要影响因素为决策者与患者之间的关系(r)以及对病情的估计(e)。

为了研究方便,把决策者与患者之间的关系分为两类:(1)患者自己为决策者;(2)决策者是患者家属。 $r \in [0, 1]$,如果患者就是决策者的话,取 $r = 0.5$,关系越亲密, r 取值越大,反之越小。

对病情估计值(e)也设置为: $e \in (0, 1]$ (此处 e 不为 0,如果为 0 意味着没有病,则无须预约专家看病),估计越严重,则越趋向于 1。对病情估计越严重,能容忍的等待时间 t_2 越短,因为此处研究的是门诊预约问题,并非急诊,所以能容忍的等待时间 t_2 的取值大于 0。 t_2 的单位根据研究问题划分时间粒度决定,此处设其单位为小时。所以有 $t_2 e = 1, e = 1/t_2$,可以推算出 $t_2 \in [1, \infty)$ 。焦虑程度为

$$A(r, e) = (1+r) \times (1+e) = (1+r) \times \left(1 + \frac{1}{t_2}\right) \quad (2.3.5)$$

当焦虑程度 $A(r, e)$ 和关系 r 两个参数已知时, 可以推算出容忍的等待时间 t_2 , 此时对应的是患者能容忍的最大等待时间 $t_{2\max}$ 。

2.3.3 模型建立

三级甲等医院的专家门诊尤其是口碑好的专家很难预约到, 但为什么很多患者或患者家属还是坚持要找到三级甲等医院找专家看病, 而不愿意选择离家近的, 或者不是那么很热门的专家看病呢? 以下对专家门诊预约中的行为决策建模分析。

1. 患者的行为决策模型

患者是否选择某个专家进行预约, 是根据他能获得的服务价值决定的。服务价值是顾客的感知利得与感知利失间的差^[33-34]。此处的感知利得是指所能获得的服务质量 $a_1 Q(z)$, 感知利失是指排队成本 T_c 。此外, 患者的焦虑程度将对服务价值产生影响, 所以综合 3 个因素, 服务价值 V 可以表示为

$$V = a_1 Q(z) - T_c - a_2 A(r, e) \quad (2.3.6)$$

式中, a_1 表示每一单位服务质量价值系数, a_2 表示每一单位焦虑程度折合的价值系数。

用 $s(i)$ 表示第 i 个患者的策略选择, $s(i) = 0$ 表示离开队列, $s(i) = 1$ 表示留在队列。显然有

$$V_i = \begin{cases} a_1 Q(z) - T_c - a_2 A(r, e) \leq 0, & s(i) = 0 \\ a_1 Q(z) - T_c - a_2 A(r, e) > 0, & s(i) = 1 \end{cases} \quad (2.3.7)$$

当服务价值大于零时, 患者会选择排队等待; 而服务价值小于零时, 患者会放弃排队, 而选择其他专家。也就是说患者只有当 $V(i) > 0$ 时才会进入队列。令 $V(i) = 0$, 此时对应的 t_2 是患者 i 能容忍的最长等待时间 ($t_{2\max}$)。将式 (2.3.1)、式 (2.3.3)~式 (2.3.5) 代入式 (2.3.7), 则有

$$V_i(t_2) = a_1(Q_b - z\mu)\alpha - \frac{c\omega_1}{\mu - \lambda} - \omega_2 p(z_{\max} + 1 - z) - \omega_3 b t_2 - a_2(1+r)\left(1 + \frac{1}{t_2}\right) = 0 \quad (2.3.8)$$

假设除了 t_2 以外的其他变量均为已知, 则求解式 (2.3.8) 可以获得 $t_{2\max}$ 的取值。

根据前述分析, 当患者的焦虑程度和关系 r 已知时, 则可计算出 $t_{2\max}$ 的值, 代入式 (2.3.8), 可推算出患者希望看诊的专家等级。当患者的焦虑程度未知时, 可以根据他预约的专家等级等信息推算出他的最大容忍等待时间 $t_{2\max}$ 。

令 n 为队列中排队等待的人数, μ 为服务率, 在 $M/M/1$ 排队系统中, 患者的预期等待时间为

$$t_2 = \frac{n+1}{\mu} \quad (2.3.9)$$

综合式(2.3.8)、式(2.3.9), 设 N_q 为患者能容忍的最长队列, 则有 $N_q = \mu t_{2\max} - 1$ 。

当患者进入预约系统, 看到某个专家的排队人数 $n \leq N_q$ 时, 他会选择进入队列排队等候, 否则, 将会选择其他专家。

如果将一个专家看作是一个服务台, 则一个专家的排队系统是 $M/M/1/K$ 类型的^[35], 其中 K 是每个时间段专家能看诊的患者数, 即为该排队系统的容量。患者到来过程和服务过程是独立的:

$$K = \overline{N_q} + 1 = \mu \overline{t_{2\max}} \quad (2.3.10)$$

式中, $\overline{N_q}$ 是所有选择该专家的患者平均可容忍最大队长, $\overline{t_{2\max}}$ 表示患者平均可容忍最长等待时间。可见 K 是 μ 的函数, 如果设定 K 为固定值, 则 μ 越小, $\overline{t_{2\max}}$ 越大。而 $\mu = \frac{1}{t_1}$, μ 越小, 意味着服务时间 t_1 越大, 患者感知的服务质量越高。 $t_{2\max}$ 越大, 意味着患者平均容忍等待时间越长。这也和现实非常吻合: 患者愿意花时间去等待服务质量高的专家。

2. 医院的应对策略分析

在专家预约排队系统里, 目标函数设置不能仅考虑医院的收益最大化, 而应该从资源优化利用、患者的满意度等角度出发。

如果把专家单位时间段内可用于看诊的时间看作是资源的话, 如何充分利用这些时间接诊更多有需要的患者将成为医院的一个非常重要的目标。

1) K 值固定时最优服务率的确定

目前大多数专家门诊每天的看诊人数是设置固定上限的, 即 K 值是固定的。

服务率 μ 由患者平均看诊时间 t_1 决定, 每个患者看诊时间 t_1 将决定患者的感知服务质量, 如果要提高患者满意度, 就应该保证每个患者的看诊时间, 设置最小看诊时间 $t_{1\min}$, 保证每个患者和专家之间的充分沟通。假设每个专家每天看诊总的时间为 T , 则有

$$t_{1\min} = \frac{T}{K} - \epsilon \quad (2.3.11)$$

式中, ϵ 是两个患者之间切换的时间和医生休息的时间。由最小看诊时间 $t_{1\min}$, 可以对应获得最大服务率 μ , 即 $\mu_{\max} = \frac{1}{t_{1\min}}$ 。

在实际操作中,可以预测每个患者就诊开始时间,提前告知患者,患者可以按时来就诊,不需花费太长时间的排队等待,从而改善患者就诊体验,提高患者满意度。这种方式适合用在患者排队等待量不大的科室或专家门诊。

2) K 值变动设置

虽然目前大部分医院专家门诊都设立每日固定号源数,但对于某些热门科室或热门专家,显然不是最合理的。主要原因有两个方面:一是每个患者看诊时间不一致,因为每个患者病情不一样,所以看诊的时间有的长,有的短,如果看诊时间比较长的患者比较多,则导致专家要加班才能完成当天的工作;二是对于热门科室或专家,号源数总是不够用的,必须要优化专家资源的利用率,才能既有效利用资源又增加患者的满意度。

假设患者的到来是参数为 λ 的泊松流,即顾客相继到达的时间间隔序列独立,服从分布 $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$ 。每个顾客所需要的服务时间也是独立的,服从参数为 μ 的分布 $F(t) = 1 - e^{-\mu t}$, $t \geq 0$ 。将一个专家看作是一个服务台,则其每个时间段的号源数可以看作系统容量。且到达过程与服务过程相互独立。

当患者挂号时,如果 K 个号源全部预约完毕,则患者不得不放弃,无法进入排队;当号源数还有余额时,患者进入队列排队。该过程可以看作是一个有限状态的生灭过程,根据其稳定状态求解公式可以推算系统中有 j 个患者的稳定概率为

$$P_j = \begin{cases} \frac{(1-\rho)\rho^j}{1-\rho^{K+1}}, & \rho \neq 1, 0 \leq j \leq K \\ \frac{1}{K+1}, & \rho = 1, 0 \leq j \leq K \end{cases}, \quad \rho = \frac{\lambda}{\mu} \quad (2.3.12)$$

对于 $M/M/1/K$ 的专家门诊预约排队系统,患者不能拿到预约号概率为 P_K ,能拿到预约号的概率为 $1 - P_K$ 。因此单位时间内能进入排队队列的患者人数为: $\bar{\lambda}_i = \lambda(1 - P_K)$ 。

假设平均每个患者能给医院带来的收益为 h ,而为每个患者服务时,医院花费的平均成本为 c_h ,则单位时间内医院的收益为

$$f(\mu) = \lambda(1 - P_K)h - c_h\mu \quad (2.3.13)$$

假设 λ, h, c_h 均为已知,由式(2.3.10)有 $K = \mu t_{2\max}$,代入式(2.3.13),由函数 $f(\mu)$ 对 μ 求导,令 $f'(\mu)$ 等于 0,可求出最优服务率 μ^* 。

$$f'(\mu) = -\lambda h P'_K - c_h = 0 \quad (2.3.14)$$

对式(2.3.14)求解,得到最优服务率后,可进一步推算出最优服务率下 K 的取值,从而可以动态设定每日号源量。

2.3.4 算例

1. 患者排队决策算例

假设某三级甲等医院规定:专家等级分 5 个层次,分别用 $z=1,2,3,4,5$ 表示,数字越小等级越高;根据上海现行医院政策规定,对专家门诊实行限号,一个专家半天只能看 30~35 个号,特殊情况下才允许加号。一个专家一天最多可以看 70 个号,每天按坐诊 8h 来算,则每个患者最小服务时间为 $\frac{8}{70}$ h,对应的最大服务率 $\mu_b=8.75$;假设某个科室规定专家一天最多可以看诊 50 个号,平均每个患者服务时间为 $t_1=\frac{8}{50}=0.16$ h。

假设目前有一个患者,需要为自己预约挂号, $r=0.5$,病情估计值为 0.8,所以可以推算出最大容忍等待时间为 $t_{2\max}=\frac{1}{0.8}=1.25$ h。若服务率、专家等级和口碑都取到最佳值,则可获得最好服务质量: $Q_b=8.75\times\frac{5}{1}\times 1=43.75$ 。假设对于该患者等待成本、专家门诊价格和沉没成本三者的权重一样,即 $w_1=w_2=w_3=\frac{1}{3}$ 。2019 年上海三级甲等医院门诊挂号费为 22 元,故设定 $p=22$,所有专家口碑都设定为 $a=1$ 。根据前文分析,患者平均到来率必须小于服务率,现假设该科室患者到来率 $\lambda=6$ 。当时上海最低工资标准为 2420 元/月,按每天工作 8h(小时),工作 22d(天)计算,折合成每小时 13.75 元,假设该患者单位时间等待成本为 $c=b=15$ 元,对患者来说服务质量和焦虑程度的价值系数分别为: $a_1=1.5, a_2=1.5$ 。

将以上数值代入式(2.3.8),则有

$$1.5 \times \left(43.75 - z \times \frac{50}{8} \right) \times 1 - \frac{1}{3} \times \frac{15}{\frac{50}{8} - 6} - \frac{1}{3} \times 22(6 - z) - \frac{1}{3} \times 15 \times 1.25 - 1.5 \times (1 + 0.5)(1 + 0.8) = 0$$

可以推算出 $z=2.889$,即这个患者将希望预约到一个等级在 2~3 的专家。

如果患者焦虑程度未知,但发现他在预约一个等级为 1 的专家,其他条件同以上假设。则可以推算出他的最大容忍时间为: $t_{2\max}=0.39$ h。

根据前面分析, $N_q=\mu t_{2\max}-1=\frac{50}{8}\times 0.39-1=1.4375$,所以当该患者看到队列里的人数小于 2 会选择进入队列,否则会选择别的专家。

2. 医院应对策略算例

当每日给定号源数 K 一定时,可以预测出每个预约患者接受服务的时间,从而提前告知患者,减少患者等待时间,目前许多医院已经实现了这种管理方式。

以下讨论根据最优服务率设定 K 值的算例。

接上例的假设, $\lambda = 4, h = 10, c_h = 5$,另设患者平均最大容忍等待时间为: $\overline{t_{2\max}} = 1\text{h}$,则 $K = \mu$ 。将以上数据代入式(2.3.13)、式(2.3.14),则有

$$f(\mu) = 40(1 - P_K) - 5\mu \quad (2.3.15)$$

$$f'(\mu) = -40P'_K - 5 = 0 \quad (2.3.16)$$

图 2-4 显示了 $f'(\mu)$ 的函数曲线,在图中 $f'(\mu) = 0$ 的点有两个,其中一个接近 0,舍弃。另外一个,使用 MATLAB 求解方程,得到 $\mu^* \approx 4.26$ 。将其代入式(2.3.15),求得 $f(\mu) = 12.0792$ 。根据前文分析,此处 $K = \mu = 4.26$,即在患者最大容忍等待时间为 1h 时,医院能提供的预约号数为 5(4.26 向上取整数)个,由此推算,一天按照 8h 工作时间算,则能提供最大服务为 $5 \times 8 = 40$ 个预约号。

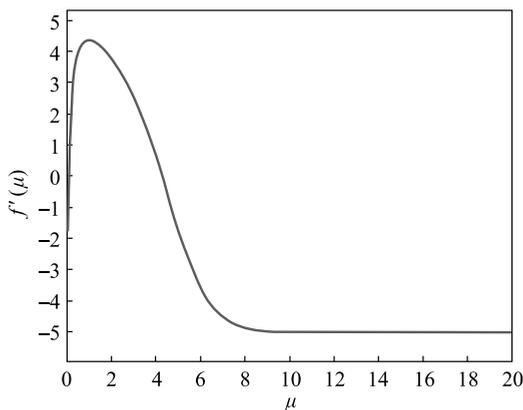


图 2-4 $f'(\mu)$ 的曲线

2.3.5 管理启示

医疗服务需求的持续增长给各级医疗机构带来了挑战。虽然目前我国已经建立了各级医疗机构,并布局到村、镇等,尽量接近居民生活地,但由于医疗资源的级别、医疗水平、患者对医疗机构的看法和对自身病情估计等因素影响,导致患者大量涌入大城市的三级甲等医院,使得这些医院不堪重负,而二级及以下医疗机构门可罗雀。在这样的背景下,研究患者在专家门诊预约中行为决

策的特征对制定医疗分诊策略变得非常具有实际意义。本部分首先对患者预约排队行为进行分析,发现患者对病情的估计和对医疗服务的期望质量判断是两个影响患者决策的重要因素;接着,分别对影响患者行为决策的主要因素进行分析,修改了服务质量模型,增加了口碑因子,以往的行为决策模型只包含服务质量和排队成本因素,在此基础上,增加焦虑程度因素,构建了包含医疗服务质量、排队成本和焦虑程度的服务价值函数分析患者行为模式,建立患者预约行为决策模型,并分析医院的应对策略。最后设计了两个算例,分别验证模型的可行性。算例1假设已知患者当前的病情估计值和焦虑程度价值系数,可以通过模型推算其期望看诊专家的级别以及他的最长容忍排队时间;算例2在患者最长容忍时间已知情况下进一步预测医院应该设置的号源数。

根据以上研究,给医院提出以下管理建议:

- (1) 医院可以设置最短看诊时间,以保证医院的利益和患者的满意度;
- (2) 每个科室可以统计和计算患者的最长容忍时间和最长容忍队长,及时调整专家数和专家看诊数量,以提高患者的满意度;
- (3) 专家号源数固定的科室或专家可以根据最优服务率调整患者看诊时间;
- (4) 专家号源数可变的科室,设置合理的专家号源数以充分利用专家资源,同时达到患者分流的目的。

患者选择医疗服务机构和医生时影响其决策主要因素是病情及其带来的焦虑心理,本部分针对这些因素进行讨论,研究专家门诊预约中患者的行为决策,并建立模型,提出求解方法。但这只是一个开始,患者行为对门诊管理的影响这个领域还存在许多问题值得探索:

一方面,患者行为决策受到许多复杂因素的影响,需要搜集大量数据进行更进一步地挖掘和分析;另一方面,基于患者的择医行为研究如何设置合理的医疗分诊机制。

2.4 门诊服务调度问题

2.4.1 相关研究概述

作为“看病难”最突出的部门,其服务质量直接决定着医院的服务水平,门诊资源的利用效率直接影响着后续部门的运作效率。为了保障门诊医疗活动能高效地开展,门诊服务管理需要将多种资源进行合理的调度,以满足不断涌入的患者需求。门诊服务管理中面临多种随机因素,包括不同病症的患者、患者不同行为、医生不同行为等,从而使得门诊调度变得非常复杂。

随着互联网技术的迅速发展和国家对“看病难”问题的重视,出台了一系列政策,门诊预约服务在全国各地得到了广泛的推行并逐渐完善,预约挂号服务已经成为医院提高门诊资源利用效率、解决门诊服务难题的重要切入点,门诊预约调度也成为医疗运作管理的一个研究热点。目前门诊预约调度研究主要从以下几个角度切入:

1. 服务时间不确定

实际操作中,由于患者服务时间的随机性,会导致预先为患者分配的服务时间与患者实际接受的服务时间不一致。服务时间的不确定性不仅增加了门诊预约调度难度,也在一定程度上影响了门诊的服务质量和资源的利用效率。

Wang^[36]指出当患者服务时间为独立同分布的指数变量时,其最优预约时间间隔服从先增加并维持在一个较高水平然后逐渐递减的穹顶形(又称为dome型)模式;Denton和Gupta^[37]基于两阶段动态线性规划模型求解门诊预约中患者的最优时间间隔,并证明了dome型预约规则会大幅提高门诊服务效率;Begen和Queyranne^[38]假设服务时间的联合离散概率分布已知且服务时间是相关的,得出了最优的门诊调度策略;Denton等^[39]指出当存在两种不同类型的患者且服务时间独立时,按照服务时间方差递增的服务次序最优。

王珊珊等^[40]考虑了门诊服务中不确定服务时间,基于服务时间联合概率分布的支撑集和矩等部分信息,并利用平均绝对偏差刻画服务时间的相关性,使最坏情况下期望等待成本和加班成本最小化,建立了分布式鲁棒优化门诊预约调度和排程模型。张文思等^[41]考虑随机服务时间与行为特征互不相同的异质患者,建立了随机混合整数规划模型,设计了启发式算法对多个患者预约方案和服务顺序同时进行优化。

2. 考虑患者特征

在实际的门诊预约中,提前预约的患者可能因为各种原因在就诊当天没有前往医院接受服务,即发生患者爽约(No-show)。Cayirli等^[42]称具有不同爽约率、治疗时间不同的患者为异质(heterogeneous)患者。Ho和Lau^[43]研究了患者爽约率、服务时间变动系数和单位服务时间患者数量三个变量对预约调度规则的影响;Liu等^[44]基于排队论模型,研究患者爽约率与预约-服务时间间隔的相关关系,设计了相应的超订机制。Zacharias和Pinedo^[45]研究了爽约率和等待时间成本不同的异质患者的预约调度准则, Lee等^[46]基于两类患者期望服务时间和爽约率的不同,设计了固定到达时间间隔的门诊预约调度方法。范喧竺等^[47]分析了门诊患者的偏好行为模式(就诊时间的偏好和不同类型门诊服务

的偏好)和患者等待期的有限耐心行为模式,采用预算分配多目标计算和遗传算法相结合的仿真优化算法,得到了两类医生(专家门诊和普通门诊)的患者联合调度方案。苏惠莽等^[48]考虑患者和医生的异质性、患者到达和服务过程的不确定性,以最小化等待时间和延迟惩罚为目标,解决在线随机调度问题,采用马尔可夫决策过程,提出了一种求解算法。

2.4.2 问题描述与分析

在门诊调度中考虑患者对时间段的选择和不同类型医生资源的选择,包含了非预约患者的到达和复杂的患者流带来的不确定的服务时间,构建的问题是NP难的。

我国目前大部分医院设置专家门诊和普通门诊,在挂号时,大部分患者选择专家门诊,尽管专家门诊的平均等待时间要长一些^[46]。而当患者发现等待时间太长不耐烦时,又退号重新挂号,进入新的队列,这样导致患者在队列中无效流动,带来低效率和患者的低满意度。

门诊每天接待的患者及其行为具有异质特征,同时医生也是有差异的,已有的成果主要围绕以下几个方面的特征展开研究:

- (1)服务时间的不同(不确定或随机性);
- (2)患者的爽约行为,部分研究学者考虑随机服务时间和患者爽约行为建立模型并找到求解方法^[41];
- (3)患者对时间段的偏好、对不同级别医生的偏好行为。

这一类型的研究主要针对患者的不同(如病症的复杂程度,对不同级别医生的偏好等)和医生的不同(如专家门诊和普通门诊,或高级医生和初级医生等)的特征,将其纳入门诊调度方案中。

一种方法是将医生分为:初级医生和高级医生两类,将患者分为正常患者和复杂患者两类。患者有两种选择:一是患者可以自由选择他们喜欢的医生,大多患者会选择高级医生;二是让系统将他们分配给任何医生,一般系统会把患者分配给等待时间最短的医生。这两种策略都会带来资源浪费或资源不足。在研究中考考虑两个方面成本:等待成本和延期成本,描述患者的到达过程和医生服务过程,通过实验可以证明患者到达过程服从泊松分布,医生服务时间服从指数分布。苏惠莽等^[48]考虑患者对初级医生和高级医生选择偏好的在线调度问题,讨论三种策略:患者自由选择他们喜欢的医生,称为自由选择策略;系统将患者分配给等待时间最短的医生,称为短视策略;采用马尔可夫决策过程(Markov decision process, MDP)代替传统的机器调度理论和仿真方法,并设计了策略迭代算法(PI算法)求解该模型,称为MDP策略。

另外,患者对不同就诊时间段和不同职称的医生有明显的偏好——更多的患者愿意选择专家门诊而非普通门诊,导致选择专家门诊的患者平均等待时间比选择普通门诊的患者长,等待时间不均导致部分患者不耐烦退出队列重新选择医生,甚至放弃继续排队,离开医院。部分患者退出队列,不仅浪费了门诊挂号资源,也意味着患者无法在有效时间内就医。范喧竹等^[47]考虑患者对专家门诊和普通门诊选择偏好及排队等待的耐心,采用数据驱动的离散事件模拟(discrete events simulation, DES)方法来模拟门诊场景,进行仿真优化和多目标优化,实现了专家门诊和普通门诊之间的帕累托(Pareto)最优联合容量规划和预约调度。

因此,考虑患者的非均衡选择行为、合理配置门诊资源、提高门诊服务效率、减少患者的不满意是门诊调度的重要问题。以下着重介绍基于MDP的在线调度方法和基于多目标优化的调度方法。

2.4.3 基于MDP的在线调度方法^[48]

1. 模型建立

从某三级甲等医院妇科电子病历系统获得2015年10月到2016年9月一年的就诊记录8万多条,包含时间段内每次访问的注册时间和服务开始时间。经过检验,确定患者到达过程遵循泊松分布,服务时间比较接近指数分布。

符号说明:

D_J : 初级医生; D_S : 高级医生;

P_G : 普通患者; P_C : 复杂患者;

n_J 表示等待医生 D_J 的患者数;

n_S 表示等待医生 D_S 的患者数;

\bar{n}_J 表示 n_J 的边界;

\bar{n}_S 表示 n_S 的边界;

S 表示状态集,其中 $S = \{(n_J, n_S) | 0 \leq n_J \leq \bar{n}_J, 0 \leq n_S \leq \bar{n}_S\}$;

$p(s_i, s_j | a(s_i))$ 为给定当前状态 s_i 的下一状态为 s_j 的概率;

$\mathbf{P}(a) = \{p(s_i, s_j | a(s_i))\}$ 对于所有 $s_i, s_j \in S$ 为策略 a 下的转移矩阵;

$\tau(s | a)$ 表示在策略 a 下从一个状态 s 到另一状态的期望时间间隔。

采用MDP解决该调度问题,首先定义5个关键元素:

(1) 决策时期: 当普通患者 P_G 来服务时就要做出决策。由于患者到达时间是一个泊松过程,因此决策时间点不是固定的,这意味着MDP模型是一个连续时间的MDP模型。

(2) 状态: MDP模型中的状态(用 $s \in S$ 表示)是指系统中的患者数。用 $n_J(n_S)$

表示等待医生 $D_j(D_S)$ 的患者数, $s=(n_j, n_S)$ 。 $n_j(n_S)$ 的边界定义为系统的能力, 用 $\bar{n}_j(\bar{n}_S)$ 表示, 从而状态集 $S=\{(n_j, n_S) | 0 \leq n_j \leq \bar{n}_j, 0 \leq n_S \leq \bar{n}_S\}$ 。

(3) 行为: 在每个决策时段, 一个行为可以决定一个普通患者 P_G 分配给 D_S 与否。 $a=\{a(s) | s \in S\}$ 表示一个随机策略, 映射每个状态到行为的分布, 即, $a(s)$ 是在状态 s 下 P_G 分配给 D_S 的可能性。用 A 表示所有可能策略的集合, 即 $A=[0, 1]^{|S|}$ 。

(4) 转移概率: 每当患者加入系统或在服务结束后离开时, 状态 s 改变。改变状态 s 的转移取决于发生的事件。设 $p(s_i, s_j | a(s_i))$ 为给定当前状态 s_i 的下一状态为 s_j 的概率。将策略 a 下的转移矩阵表示为 $P(a)=\{p(s_i, s_j | a(s_i))$ 对于所有 $s_i, s_j \in S\}$ 。

(5) 成本: 与状态 $s=(n_j, n_S)$ 相关的成本由两部分构成。一个是患者的等待成本, 等待服务的患者数是 $(n_G-1)^+ + (n_S-1)^+$, 他们一直等到状态改变。用 $\tau(s | a)$ 表示在策略 a 下从一个状态 s 到另一状态的期望时间间隔。 $((n_G-1)^+ + (n_S-1)^+) \times \tau(s | a)$ 是状态 s 下的期望等待时间。另一个是由于患者延期带来的惩罚成本。在边界状态, 根据政策, 即将到来的患者可以推迟到第二天。 $\rho(s | a)$ 表示在策略 a 下状态 s 时推迟一个患者的可能性。策略 a 下状态 s 时总成本为 $f(s | a)=((n_G-1)^+ + (n_S-1)^+) \times \tau(s | a) + h(\rho(s | a))$, 单位等待成本规范化后为 1, $h(\cdot)$ 为延迟成本函数。 $f(a)$ 是 $f(s | a)$ 的矢量。

令 $\pi(a)=\{\pi(s | a)$ 对所有 $s \in S\}$ 为策略 a 下的稳定状态概率。可以使用如下公式计算:

$$\pi(a)P(a) = \pi(a) \quad (2.4.1)$$

使用 $\pi(a)P(a)$ 作为策略 a 下的期望平均成本 $\eta(a)$, 因为 MDP 是遍历的。目标是要找到策略 a^* , 从而最小化期望总成本 $\eta(a^*)$, 即

$$\begin{aligned} & \min_{a \in A} \pi(a)f(a) \\ & \text{s. t. } \pi(a)P(a) = \pi(a) \end{aligned} \quad (P)$$

用 $g(s | a)$ 表示策略 a 下的状态 s 的潜在性, 定义为

$$g(s | a) = f(s | a) - \eta(a) + \sum_{s' \in S} p(s, s' | a(s))g(s' | a) \quad (2.4.2)$$

$g(s | a)$ 可以表示为对于期望平均成本的 s 状态的长期“潜在性”贡献。特别地, $f(s | a) - \eta(a)$ 是当前状态 s 的下一步贡献, $\sum_{s' \in S} p(s, s' | a)g(s' | a)$ 是下一状态的期望长期“潜在性”贡献。式(2.4.2)的矩阵形式就是泊松公式:

$$(I - P(a))g(a) + \eta(a)e = f(a) \quad (2.4.3)$$

这里 I 和 P 是相同维度的识别矩阵, e 和 f 是相同维度的单元矢量。

因为“潜在性”是一个相对值, 无法从公式获得一个唯一的 $g(a)$, 需要通过

设置 $\pi(a)g(a) = \eta(a)$ 进行规范化。将有

$$(I - P(a) + \epsilon\pi(a))g(a) = f(a) \quad (2.4.4)$$

策略迭代算法

问题(P)的策略迭代算法(PI算法)步骤如下:

输入:MDP模型

输出:一个优化策略 a^*

初始化:猜测一个初始策略 a_0 ,设置 $k=0$;

While 未达到最优 do

- ① 在策略 a_k 下计算转移概率 $P(a_k)$;
- ② 采用式(2.4.1)求解对应稳定状态概率 $\pi(a_k)$;
- ③ 通过泊松公式(2.4.4)获得潜在性 $g(a_k)$;
- ④ 求解 $\min_{a \in A} \{P(a)g(a_k) + f(a)\}$,用 a_{k+1} 表示解;

If $a_{k+1} = a_k$ then

获得最优解;

$a^* := a_{k+1}$;

返回 a^* ;

Else

$k := k + 1$

End

End

2. 算例

患者自己选择医生的策略(自由选择策略)描述如下:

$$a(s) = \begin{cases} 1, & n_s < \bar{n}_s \\ 0, & n_s = \bar{n}_s \end{cases} \quad (2.4.5)$$

由系统分配患者到医生处(短视策略)描述如下:

$$a(s) = \begin{cases} 1, & n_s < n_j \\ 0, & n_s \geq n_j \end{cases} \quad (2.4.6)$$

我们对患者的到达和离开过程进行了仿真。通过使用相同的随机数发生器种子来保证公平性。模拟时段为 8h,在这 8h 内,记录每个患者的等待时间,统计服务患者和延迟患者的数量。以每位患者的平均等候时间和延迟患者的数目作为评估标准。为了消除可变性的影响,对于每个参数设置,进行 100 次模拟并取平均值。两种类型患者的到达率范围为每秒 0.0001~0.0028,步长为 0.0003。在相同的步长下,服务率的范围为 0.0007~0.0034。使用系数在

5400~21 600 的线性函数来表示延迟成本 $h(\cdot)$ 。系数是单位延期成本。例如,系数 5400 意味着推迟一名患者的费用与让一名患者等待 5400s 的费用相同。

复杂患者到达率 λ_C 变化对患者等待时间的影响:当 λ_C 较高时,系统中的患者等待的时间较长,自由选择策略的性能总是最差的;当 λ_C 很小时,短视策略和 MDP 策略将适当数量的普通患者分配给 D_S ,自由选择策略分配更多,因此自由选择策略的性能更差。当有更多的复杂患者时,就很难依靠短视策略和 MDP 策略来改善系统。当 λ_C 较低时,短视策略与 MDP 策略的差异不显著,当 λ_C 相对较高时,MDP 策略仍优于短视策略。原因是 MDP 策略预期可能会有许多复杂的患者分配给 D_S 。

复杂患者到达率 λ_C 变化对延迟患者数的影响:当 λ_C 越高时,系统延迟的患者越多。特别是当 λ_C 较大时,自由选择策略的性能较差。当 λ_C 较小时,短视策略和 MDP 策略延期患者的数量非常接近,而当 λ_C 较大时,短视策略延期患者的数量更多。当 λ_C 很小时,系统中很少有长队列,所有策略延迟患者都较少。当 λ_C 变大时,MDP 策略将适当数量的普通患者分配给 D_S ,因此使用 MDP 策略的延迟患者最少。

普通患者的到达率 λ_G 对平均等待时间的影响:随着 λ_G 的增大,短视策略和 MDP 策略患者在系统中的等待时间先减小后增大。当 λ_G 较小时,正常患者很少等待,因此等待费用大部分来自复杂患者。随着 λ_G 的增加,总等待时间并没有增加很多,但患者的数量却在增加。这就是为什么当 λ_G 稍微增加时,每个患者的平均等待时间更短。还可以看到,当 λ_G 较高时,自由选择策略的缺点更为显著,MDP 策略的性能优于短视策略。

普通患者的到达率 λ_G 对延迟患者数的影响:自由选择策略的结果是延期患者数增加,在这种策略下,只要 D_S 可用,普通患者总是去 D_S 那里,因此,普通患者越多,延迟患者越多。对于短视策略和 MDP 策略,当 λ_G 较小时,延迟患者数略有增加,当 λ_G 较大时,延迟患者数迅速增加。其原因是,当 λ_G 不太大时,这两种策略可以通过将一些普通患者分配给 D_S 来平衡工作量。然而,当 λ_G 较大时,由于系统已经过载,影响减弱。

服务率 μ 变化对平均等待时间的影响:当 μ 增大时,所有策略下的平均等待时间都减小,但自由选择策略的平均等待时间仍然最长。与短视策略和 MDP 策略相比,当 μ 较小时,MDP 策略的性能明显优于短视策略, μ 增加后,MDP 策略的优势减小了。

服务率 μ 变化对延迟患者数的影响:在所有策略下,服务率 μ 都会降低。与其他两种策略相比,随着服务率的增加,自由选择策略先增加后减小。当服务率太高时,无论如何分配正常的患者,很少有患者可以推迟;当服务率太小时,系统高度拥挤,因此所有策略都没有多大帮助。

延迟成本对平均等待时间和延迟患者数量的影响:随着延迟费用的增加,MDP 策略下患者的等待时间增加,而自由选择策略和短视策略下患者的等待时间保持稳定。自由选择策略和短视策略不考虑延迟成本,因此,延迟患者的数量没有变化,平均等待时间也没有变化。然而,延迟成本对 MDP 策略有很大的影响。当延迟成本较大时,因为服务的患者数增加带来更大的延迟成本,导致平均等待时间延长,MDP 策略会减少延迟患者数量。即使延迟成本很大,MDP 策略下的平均等待时间比其他两种策略要短得多。

虽然可以看到 MDP 策略在平均等待时间方面比短视策略有很大的改进,但不能说 MDP 策略总是更好,因为它在某些情况下会导致更多的延迟患者。与短视策略相比,MDP 策略在不增加延迟患者数量的前提下,减少了每位患者的平均等待时间,从而降低了总成本。

经验结果表明,在某些情况下,服务时间不是指数分布的。为了充分评估 MDP 策略的价值,实现了短视策略和服务时间确定的 MDP 策略。由于在服务时间呈指数增长的情况下,自由选择策略的性能最差,因此这里不考虑它。在 30 种不同的参数设置下进行了 100 次模拟,并进行配对测试,比较短视策略和 MDP 策略在每个患者平均等待时间、延迟患者数量和总费用方面的性能差异。总成本和待检验的假设与前文相同。即使在确定的服务时间下,MDP 策略在每个服务患者的平均等待时间和总成本方面也优于短视策略,而不会增加延迟患者的数量。因此,即使服务时间是确定的,MDP 策略也比短视策略具有更好的性能。

2.4.4 基于多目标优化的调度方法^[47]

从某医院信息系统搜集获取门诊部 2016 年全年门诊数据,统计了 411 672 例患者的挂号时间、排队退出时间、开始就医时间、开始结算时间、处理时间等时间戳,其中 26 856 例患者未就诊就离开。对两类门诊的患者选择行为、平均等待时间等数据进行描述性统计分析,结果如表 2-2 所示。

表 2-2 描述统计分析和分布拟合^[48]

对比项目	普通门诊	专家门诊
患者选择/%	37	63
平均等待时间/min	19.55	33.91
平均服务时间/min	4.01	4.84
服务时间分布	$0.5 + \text{LOGN}(2.56, 2.67)$	$0.5 + \text{LOGN}(2.91, 3.69)$
到达间隔分布	$-0.001 + \text{EXPO}(1.56)$	—

表 2-2 中数据显示选择专家门诊的比例明显比普通门诊要高,且平均等待时间要长得多,而平均服务时间虽然长一点,但仅多了 0.83min。同时,统计数据也发现专家门诊排队人数多,比较拥挤。同时放弃排队的患者数也在繁忙时段出现高峰。定义患者放弃前的等待时间为患者的等待耐心阈值,用 WPT (waiting patience threshold)表示。患者针对候诊室的拥挤程度和其他患者的平均候诊时间形成自己候诊时间的预期值,将同一时间段内其他患者在候诊室的平均等待时间定义为患者等待时间期望值 WTE(waiting time expectation)。

对于时间段 i , 对应的患者等待耐心阈值为:
$$WPT_i = \frac{\sum_i \text{离开前的等待时间}}{\sum_i \text{离开的患者数}},$$

患者等待时间期望值为:
$$WTE_i = \frac{\sum_i \text{服务前的等待时间}}{\sum_i \text{服务的患者数}}。$$
 等待时间期望值增

加将使等待耐心阈值增加,可表示为

$$\frac{\partial WPT_i}{\partial WTE_i} > 0 \quad (2.4.7)$$

定义回归模型:

$$WPT_i = \beta_0 + \beta_1 \times WTE_i + \epsilon_i \quad (2.4.8)$$

式中, ϵ_i 为平均零误差项。 β_1 表示使用时间相对于荷载的弹性。 $\beta_1 > 0$ 时表示患者通过增加等待耐心阈值对更长的等待时间期望值作出反应,从而支持公式(2.4.7)。根据回归方程,可以利用仿真模型中患者的实际平均等待时间估计患者的耐心阈值来判断其行为模式,患者放弃率可以作为衡量系统效率和患者满意度的定量评价指标。

1. 模型建立

使用 DES 工具 Arena 建立模拟模型,模型考虑了以下要素:门诊患者的比例和到达间隔分布、两个医生队列之间的基本患者流、患者对医生服务的选择(专家门诊和普通门诊)、患者等待的行为模式、不同类型医生的耐心和服务时间分布。模型按照以下流程进行模拟:

人数设置:假设平均每天有 1400 名患者,大多数患者都有预约,而其他患者则是当天到达的未预约患者。

在进入候诊室之前,患者被分为两组:

(1) 预约患者,他们使用在线预约系统选择医生和预约时间。在到达诊所之前,他们有一定的概率会取消预约或爽约,其余预约的患者会准时到达门诊部,直接进入候诊室,不用再挂号。

(2)未预约而进入门诊部的患者,需在挂号窗口挂号,待选定医生有剩余容量时,方可进入候诊室。

如果医生排队等候的患者总数未达到等候室容量限制,两类患者进入等候室,按“先到先得”的顺序排队等候诊断。

进入候诊室后,患者会估计等待时间,并根据拥挤的情况调整耐心阈值。当等待时间超过患者的耐心阈值而他们不在队列的最前面时,患者将放弃在队列中等待。

当患者放弃排队等候时,有些患者选择直接离开医院。其余患者重新选择其他医生排队,重新登记,重新进入候诊室。

将模拟模型的关键绩效指标与实际数据进行比较,如接受普通门诊和专家门诊的患者的等待时间、放弃普通门诊或专家门诊队列的患者数量、医生的服务时间和利用率。表 2-3 表明,模拟结果与实际数据之间没有显著差异。

表 2-3 仿真模型验证^[47]

对比项目	门诊类别	实际数据	平均模拟结果	错误率/%
等待时间	普通门诊	19.55min	21.12min	8.031
	专家门诊	33.91min	35.37min	4.306
服务时间	普通门诊	4.01min	4.11min	2.494
	专家门诊	4.84min	4.72min	-2.479
利用率	普通门诊	74.44%	71.58%	-3.842
	专家门诊	97.78%	99.70%	1.964
离开的患者	普通门诊	1.5 人	1.6 人	6.67
	专家门诊	5.67 人	5.2 人	-8.29
放弃治疗的患者	普通门诊	3.33 人	3.1 人	-6.9
	专家门诊	22 人	22.8 人	3.64

注:离开的患者和放弃的患者人数为一天中按每小时为时间槽统计的人数平均值。

为了获得门诊患者的调度方案,建立多目标函数,分别考虑专家门诊(EOS)和普通门诊(GOS)的加权资源利用率和未就诊患者数量及医院效益:

$$\begin{aligned} \max B_H = & \mu_{EOS} U_{EOS}(S_{EOS}, S_{GOS}) + \\ & \mu_{GOS} U_{GOS}(S_{EOS}, S_{GOS}) \end{aligned} \quad (2.4.9)$$

$$\begin{aligned} \min N_{LWBS} = & \epsilon_r R(S_{EOS}, S_{GOS}) + \epsilon_b B(S_{EOS}, S_{GOS}) + \\ & \epsilon_c C(S_{EOS}, S_{GOS}) \end{aligned} \quad (2.4.10)$$

符号说明如下:

B_H : 医院收益, 两类医生在工作期间的有效服务时间百分比的加权平均数;

μ_{EOS} : 一个专家门诊医生为医院创造的单位价值, 用于表示医生的重要程度;

U_{EOS} : 专家门诊医生有效服务时间百分比;

μ_{GOS} : 一个普通门诊医生为医院创造的单位价值, 用于表示医生的重要程度;

U_{GOS} : 普通门诊医生有效服务时间百分比;

N_{LWBS} : 不可见的离开(left without being seen)人数, 包括被拒绝、错过和
中途更换医生的患者人数总和;

R : 被拒绝的患者人数;

B : 错过就诊的患者人数;

C : 中途更换医生的患者人数;

ϵ_r, ϵ_b 和 ϵ_c : 分别表示被拒绝、错过和中途更换医生的患者不满意程度的权重参数;

S_{EOS}, S_{GOS} : 是表示预约专家门诊和普通门诊服务的患者的联合调度方案的决策变量。

由于目标冲突, 多目标优化问题往往没有满足所有优化目标的最优解, 在这个问题中, 由于绩效指标(目标)的真实分布以及其他信息的均值和方差分布都是未知的。通过反复模拟, 利用观测样本来估计输出的真实分布, 然后根据估计的分布计算相关参数。针对仿真计算量大的问题, 引入预算分配多目标计算(multi-objective computing of budget allocation, MOCBA)算法来确定仿真副本的最优分配, 使获得非占优帕累托集的概率最大, 而计算量最小。引入遗传算法(genetic algorithm, GA)观测帕累托集的更新, 直到满足终止条件, 算法步骤如下:

步骤 1 初始化: 设置迭代 $v=0$ 和 $N_1^v, N_2^v, \dots, N_n^v = N_0$, 构造观察帕累托集 $S_P^{\text{obs}} = \phi, \text{Gen } g=0$ 。

步骤 2 评估性能, 并对每个设计观察到的帕累托集运行 N_0 次复制: 计算 NDP_i 并更新观察的帕累托集, 计算新的分配。对设计 $i=1, 2, \dots, n$ 按照确定总数 $\Delta = \min(\tau, \max(0, N_i^{v+1} - N_i^v))$ 提高模拟复制数, 设置 $v = v + 1$, 如果观察的帕累托集没有提升, $v \geq v_{\max}$, 复制过程结束, 否则, 返回步骤 2。如果 $g \geq g_{\max}$, 优化过程结束。否则转步骤 3。

步骤 3 子代更新: 通过观察的帕累托方案集, 基于 NDP 对集合 $\Pi = \text{SortPop}(\cdot)$ 进行排序, 并选择顶部项目作为父项集。设置染色体交叉率为 δ , 根据变异概率 P_m 和多项式变异操作符 η 改变子代。更新子代 $x = \text{PolyMutation}(x, \eta, P_m)$, 设置 $g = g + 1$, 返回步骤 2。

2. 算例

算例包含三个实验：

实验一，通过比较相关算法的效率，验证算法的有效性；

实验二，提出优化调度方案，通过对优化前后系统性能指标的比较，证明提出的优化决策支持框架能够显著提高系统关键指标的性能；

实验三，对关键参数进行实验分析，为医疗管理机构提供政策建议。

以下分别介绍三个实验过程：

1) 验证算法有效性

首先随机产生第一代解集，包括 50 个可行的调度方案，然后分别用 MOCBA 和统一计算预算分配(uniform computing budget allocation, UCBA)对其进行优化和迭代，并比较了两种算法运行中的两种错误类型：I 型错误和 II 型错误。考虑到有限的仿真预算，设定了最大的仿真复制次数，并在 MOCBA 和 UCBA 中估计了 500 次优化复制的 I 型误差和 II 型误差。在 UCBA 中，300 次重复后，两种类型的估计误差分别为 0.175 和 0.214；在 MOCBA 中，300 次重复后，两种类型的估计误差分别为 0.105 和 0.123。与 UCBA 相比，MOCBA 在相同的计算预算内，大大提高了迭代过程的精度。

2) 提出优化调度方案

从帕累托前沿解集中选取一组作为推荐的调度方案。预约患者的排程安排，在考虑到许多实际情况时，如患者的不耐烦和等待过程中的放弃行为等，和前文研究及以往文献中提到的标准的穹顶模型有所差异。当考虑到在等待过程中放弃的患者和临时增加患者时，调度方案可以采用时段过剩和不足的交替模式。为了避免预约患者过于偏向专家门诊，预约方案中尽量增加普通门诊的比例。

表 2-4 根据两类门诊的四个绩效指标，将推荐的调度方案与当前方案进行比较。表中还说明了两种医生联合调度方案带来的显著改进。除专家门诊患者的利用率外，其他绩效指标也有所提高，其中普通门诊的利用率没有大幅度降低，专家门诊医生的利用率显著提高，同时专家门诊患者的 N_{LWBS} 也显著降低，说明专家门诊和普通门诊之间，患者排班更加均衡，同时系统的整体效率得到显著提高。

表 2-4 优化方案与基础方案比较^[47]

对象	基础方案	建议方案	提高比例/%
使用专家门诊	99.76%	98.53%	-1.23
使用普通门诊	71.07%	95.32%	34.12

续表

对象	基础方案	建议方案	提高比例/%
$N_{L,WBS}$ -普通门诊	7.36 人	5.24 人	28.80
$N_{L,WBS}$ -专家门诊	27.15 人	14.38 人	47.03

3) 参数分析

对门诊系统中的关键参数,如患者对医生的偏好概率、患者的等待耐心容忍度参数、患者对医生参数变化的接受度等进行模拟实验分析,通过对这些关键参数的分析测试,可以总结出对门诊医疗管理的启示和建议。

(1) 患者偏好医生的概率:随着患者对医生偏好概率的增加,在现有的参数配置下,系统的性能会变差。这是因为患者对这两类门诊服务的偏好差异较大,意味着这两类资源无法整合利用的可能性较大,而相应的服务效率和资源利用率较低。

(2) 患者接受改变医生参数:随着患者更换医生资源的时间成本增加,系统性能显著提高,表明当两种资源之间的“共享壁垒”被打破时,门诊的医疗资源可以更加灵活。共享使得在一个资源拥挤时使用另一个资源变得容易。

2.4.5 管理启示

考虑服务时间的不确定性时,将患者分为初诊患者和复诊患者两种类型,当门诊的服务次序固定时,同种类型患者的预约服务时长总体呈现出递减的趋势,最优的服务次序与医生、患者的相对时间成本和服务时间的相关程度有关。医院管理者可根据实际情况,按照以下原则选择最优的门诊预约服务次序:

当相对时间成本适中、相关性较强时,一般将复诊患者安排在序列的前面和后面,复诊患者的平均服务时间较短,所以复诊患者的预约服务时长一般也比初诊患者的预约服务时长短,此时的最优预约服务时长呈现先增后减的趋势。

当相对时间成本较高或服务时间相关程度较大时,应适当增大序列后面的患者预约服务时长,减小序列前面的患者预约服务时长。

在排程模型中,不考虑服务时间相关性的平均成本要高于考虑服务时间相关性的平均成本。

当患者服务时间为独立同分布的随机变量时,分配给患者的服务时间呈现先增加并维持在一个较高水平,后逐渐减少的“穹顶”形,即给较早接受服务的患者和最后接受服务的患者预留较短的服务时间,而给服务顺序处于队列中间

的患者预留较长的服务时间；当患者服务时间分布互不相同，与基于样本均值近似方法的结果相比，启发式算法在求解效率和计算时间上都具有一定的优越性。

当患者等待时间成本较低时，根据数值结果，系统会为队列中间的患者安排较长的预约时间间隔，若当前患者在下一个患者到达前结束服务，则会产生医生空闲时间。由于医疗服务过程中，医生空闲往往伴随着设备空转、医疗资源闲置等，空闲时间成本往往较高，实际应用中，医院可考虑在医生服务能力范围内，安排部分当天到达的患者接受服务，以减少医疗资源的闲置；同时，为预约患者赋予较高的优先权，防止预约患者因等待时间过长而造成满意度下降。在对患者的服务顺序进行决策时，需要权衡患者等待时间成本和随机服务波动性对系统的影响，将服务时间波动较小、等待时间成本较高的患者指派至等待队列的前面接受服务，将服务时间波动性大且等待时间成本较小的患者安排在后面接受服务，以最小化由于患者服务时间的波动性对系统造成的影响，并减少预约调度系统总成本。

考虑患者对不同医生的偏好和医生之间的差异，医院在门诊资源调配时可以从以下几个方面进行优化：

第一，投资建立调度中心。分析表明，让患者自由选择策略的性能并不好，缺乏调度中心会导致患者等待时间变长，延迟患者数量变多。系统分配患者就诊策略能显著改善整个过程，使用调度策略所节省的成本远远超过了建立调度中心的成本。

第二，当系统不忙或服务时间较短时，资源（医生的可用时间）相对足够，能服务所有患者，因此系统分配患者就诊策略（短视策略）和 MDP 策略的绩效非常接近。然而，短视策略下，开发新的信息系统和培训员工需要花费时间和金钱，其价值却并不显著，因此在这种情况下实施“短视策略”是合适的。

第三，在发展中国家卫生保健资源不足，门诊部总是很忙，工作量很大。基于以上分析，预期系统繁忙时，短视策略的性能与 MDP 策略的性能相差较远。这意味着需要仔细设计调度策略，以实现整个系统的良好性能。MDP 策略在理论上是稳态下的最优策略，在仿真研究中，MDP 策略在各种情况下都显示出很大的优势。

通过分析门诊中专家门诊和普通门诊两类医生资源使用的不均衡性，发现患者对医生资源和治疗时间的偏好导致门诊服务效率低下。在考虑患者偏好的基础上，为患者安排预约可以改善门诊患者的状况。

医疗管理部门不应过分强调专家门诊和普通门诊两种医疗资源的区别，根

据患者的需求进行分诊。患者在病情不严重时,应主动选择普通门诊,同时应引导患者合理选择专家门诊,可显著提高门诊服务效率。

对于具有多种不同资源的服务系统,降低服务提供者对不同资源的偏好可以有效提高系统效率,而打破不同资源之间的共享壁垒可以进一步提高系统的服务效率。建议医疗管理部门完善门诊变更服务流程,简化医生资源共享流程,并加强对重新选择医生的指导,以便在不占用额外门诊资源的情况下,更容

2.5 小结

本章对门诊服务管理与优化问题进行了探讨,首先介绍门诊管理基本内容,梳理门诊管理流程,整理出门诊管理关键问题包括门诊预约问题、门诊流程优化问题、考虑患者行为的门诊资源配置和门诊服务调度等。

针对专家门诊预约管理,考虑患者和医生的偏好,采用匹配理论建立模型并提出求解方法,同时对专家门诊预约流程提出改进方案。研究表明:均衡匹配算法得到的匹配结果能较好地平衡专家特长与患者选择专家意向性,提高双方的满意度。

分析患者在门诊预约或挂号时的行为模式,考虑患者在择医时焦虑程度、对专家门诊服务质量期望值和医生的口碑等因素,建立服务质量模型,分析患者的就医成本,基于排队论建立患者择医行为决策模型,以此为基础给医院提出应对策略,即如何设置最优服务率和号源数,帮助医院实现门诊资源配置优化,并为实现患者有效分流提供一个途径。

门诊患者调度是一个非常复杂的问题,需要考虑患者接受服务时间的不确定性、患者的异质性、患者对不同级别医生和不同时间段的偏好等问题,针对以上问题分别介绍最新的研究成果,包括以下两个方面:

(1)针对服务时间不确定情形建立门诊预约调度模型,第一步根据不同类型患者的随机服务时间,确定最优的预约服务次序;第二步在门诊排程固定的前提下,确定患者最优的预约服务时长。

(2)考虑患者对不同级别医生和不同时间段的偏好,采用仿真的方法模拟门诊患者到达和医生服务过程,使用 MDP 或多目标优化建模方法探索调度方案。

以上方法均能从不同角度对门诊患者调度问题建立数学模型,并就算法进行讨论,以期能够降低患者等待成本,提高医院资源利用率,为门诊服务管理优化提供可供参考的解决方案。

参考文献

- [1] BAILEY N. Study of queues and appointment systems in hospital outpatient departments with special reference to waiting-time[J]. *Journal of the Royal Statistical Society*, 1952, 14(2):185-199.
- [2] MANLOVE D F. *Algorithmics of matching under preferences*[M]. Singapore: World Scientific Publishing Co Pte Ltd. , 2013.
- [3] CAYIRLI T, VERAL E. Outpatient scheduling in health care; a review of literature [J]. *Production & Operations Management*, 2010, 12(4): 519-549.
- [4] GUPTA D, DENTON B. Appointment scheduling in health care; challenges and opportunities[J]. *IIE transactions*, 2008, 40(9): 800-819.
- [5] WELCH J D, NORMAN T J, BAILEY M A. Appointment systems in hospital outpatient departments[J]. *The Lancet*, 1952, 259(6718): 1105-1108.
- [6] KAANDORP G C, KOOLE G. Optimal outpatient appointment scheduling[J]. *Health Care Management Science*, 2007,10(3): 217-229.
- [7] FELDMAN Z, MANDELBAUM A, MASSEY W A, et al. Staffing of time-varying queues to achieve time-stable performance[J]. *Management Science*, 2008, 54(2): 324-338.
- [8] MOORE C G, WILSON-WITHERSPOON P, PROBST J C. Time and money: effects of no-shows at a family practice residency clinic[J]. *Family Medicine*, 2001, 33(7): 522-527.
- [9] GUPTA D, WANG L. Revenue management for a primary-care clinic in the presence of patient choice[J]. *Operation Research*, 2008, 56(3): 576-592.
- [10] KLASSEN K J, ROHLEDER T R. Scheduling outpatient appointment in a dynamic environment[J]. *Journal of Operation Management*, 1996, 14(2): 83-101.
- [11] KIM S, GIACHETTI R E. A stochastic mathematical appointment overbooking model for healthcare providers to improve profits[J]. *IEEE Transactions on Systems Man, and Cybernetics-Part A: Systems and Humans*, 2006, 36(6): 1211-1219.
- [12] LAGANGA L R, LAWRENCE S R. Clinic overbooking to improve patient access and increase provider productivity[J]. *Decision Science*, 2010, 38(2): 251-276.
- [13] ZENG B, TURKCAN A, LIN J, et al. Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities[J]. *Annual Operation Research*, 2010, 178(78): 121-144.
- [14] RISING E J, BARON R, AVERILL B. A system analysis of a university health service outpatient clinic[J]. *Operations Research*, 1973, 21(5): 1030-1047.
- [15] COX T F, BIRCHALL J P, WONG H. Optimizing the queuing system for an ear, nose and throat outpatient clinic[J]. *Journal of Applied Statistics*, 1985, 12(2): 113-126.

- [16] CHUN P Y, BIN D S, HU C W. Measuring the efficiency of the outpatient process with queuing theory model[J]. *China Journal of Hosp Admin*, 2005, 12: 806-809.
- [17] ZHONG L W, BAI Y Q. Equivalence of two-sided stable matching[J]. *Journal of Combinatorial Optimization*, 2018, 36(4): 1380-1387.
- [18] FEDER T. *Stable networks and product graphs*[M]. Boston: American Mathematical Society, 1995.
- [19] 朱恒鹏. 建立分级诊疗体系缓解医患矛盾冲突[N]. *中国医药报*, 2014-03-24(6).
- [20] 吕键. 论深化医改进程中分级诊疗体系的完善[J]. *中国医院管理*, 2014, 34(6): 1-3.
- [21] 王克. “分级诊疗”, 一个美丽的神话? [J]. *决策探索*, 2014(4): 61-63.
- [22] 张春瑜, 李天庆. 大型综合性医院患者就医行为影响因素分析[J]. *卫生经济研究*, 2009(10): 32-33.
- [23] ADAMS E K, WRIGHT G E. Hospital choice of Medicare beneficiaries in a rural markets: why not the closest? [J]. *The Journal of Rural Health*, 2010, 7(2): 134-152.
- [24] BRONSTEIN J M, MORRISEY M. Bypassing rural hospitals for obstetric care[J]. *Journal of Health Politics, Policy & Law*, 1991, 16(1): 87-118.
- [25] ALI M M. Patient choice of a hospital: implications for health policy and management [J]. *International Journal of Health Care Quality Assurance*, 2014, 2: 152-164.
- [26] SMITH H, CURRIE C, CHAIWUTTISAK P, et al. Patient choice modeling: how do patients choose their hospitals? [J]. *Health Care Management Science*, 2017, 21(2): 1-10.
- [27] WALTER B. Choice in the presence of experts: the role of general practitioners in patients' hospital choice[J]. *Journal of Health Economics*, 2018, 60: 98-117.
- [28] 周华, 周水银. 基于顾客排队行为的专业服务等级的决策问题[J]. *中国管理科学*, 2014, (22)2: 85-93.
- [29] BANSAL H S, VOYER P A. Word-of-mouth processes within a services purchase decision context[J]. *Journal of Service Research*, 2000, 3(2): 166-177.
- [30] 黄孝俊, 徐伟青. 口碑传播的基本研究取向[J]. *浙江大学学报(人文社会科学版)*, 2004, 34(1): 125-132.
- [31] 李林, 孙军华, 周章金. 基于复杂社会网络的服务质量口碑传播[J]. *系统工程*, 2009, 27(6): 1-7.
- [32] ANAND K S, PAÇ M F, VEERARAGHAVAN S. Quality speed conundrum: trade-offs in customer-intensive services[J]. *Management Science*, 2011, 57(1): 40-56.
- [33] ZEITHAML V A. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence[J]. *Journal of Marketing*, 1988, 52(3): 2-22.
- [34] GRONROOS C. Value-driven relational marketing: from products to recourses and competencies[J]. *Journal of Marketing Management*, 1997, 13(5): 407-419.
- [35] 李娜, 贾博, 江志斌, 等. 考虑顾客体验的排队系统研究[J]. *工业工程与管理*, 2012, 17(3): 36-46.

-
- [36] WANG P P. Static and dynamic scheduling of customer arrivals to a single-server system[J]. *Naval Research Logistics*, 1993, 40(3): 345-360.
- [37] DENTON B, GUPTA D. A sequential bounding approach for optimal appointment scheduling[J]. *IIE Transactions*, 2003, 35(11): 1003-1016.
- [38] BEGEN M A, QUEYRANNE M. Appointment scheduling with discrete random durations[J]. *Mathematics of Operations Research*, 2011, 36(2): 240-257.
- [39] DENTON B, VIAPIANO J, VOGL A. Optimization of surgery sequencing and scheduling decisions under uncertainty[J]. *Health Care Management Science*, 2007, 10(1): 13-24.
- [40] 王珊珊, 李金林, 彭春, 等. 不确定服务时间下分布式鲁棒门诊预约调度和排程[J]. *系统工程学报*, 2019, 34(4): 566-576.
- [41] 张文思, 李金林, 冉伦, 等. 随机服务时间下异质患者门诊预约调度优化[J]. *运筹与管理*, 2020, 29(5): 26-36.
- [42] CAYIRLI T, EMRE V. Outpatient scheduling in health care: a review of literature [J]. *Production and Operations Management*, 2003, 12(4): 519-549.
- [43] HO C J, LAU H S. Evaluating the impact of operating conditions on the performance of appointment scheduling rules in service systems [J]. *European Journal of Operational Research*, 1999, 112(3): 542-553.
- [44] LIU N, ZIYA S, KULKARNI V G. Dynamic scheduling of outpatient appointments under patient no-shows and cancellations[J]. *Manufacturing & Service Operations Management*, 2010, 12(2): 347-364.
- [45] ZACHARIAS C, PINEDO M. Appointment scheduling with no-shows and overbooking[J]. *Production and Operations Management*, 2014, 23(5): 788-801.
- [46] LEE S J, HEIM G R, SRISKANDARAJAH C, et al. Outpatient appointment block scheduling under patient heterogeneity and patient no-shows [J]. *Production and Operations Management*, 2018, 27(1): 28-48.
- [47] FAN X Z, TANG J F, YAN C J, et al. Outpatient appointment scheduling problem considering patient selection behavior: data modeling and simulation optimization[J]. *Journal of Combinatorial Optimization*, 2021, 42(1): 677-699.
- [48] SU H Q, WAN G H, WANG S. Online scheduling for outpatient services with heterogeneous patients and physicians [J]. *Journal of Combinatorial Optimization*, 2019, 37(1): 123-149.

第3章 急诊管理与优化

3.1 急诊管理概述

3.1.1 急诊管理基本内容

急诊医学科(室)或急诊医学中心是医院中重症患者最集中、病种最多、抢救和管理任务最重的科室,是所有急诊患者入院治疗的必经之路。

综合医院急诊设有全科、内、外、妇、儿、五官、发热、腹泻等专科诊室。因此,急诊科的工作可以说是医院总体工作的缩影,直接反映了医院的急救医疗、护理工作质量和人员素质水平。

急诊管理基本内容包括:接收紧急就诊的各种患者;接收院外救护转送的伤病患者;负责对急诊和院外转送到急诊科的危重患者进行抢救;承担灾害、事故的急救工作;开展急救护理的科研和培训;建立健全急诊人员的岗位职责、规章制度和技术操作规范,培训急诊医生和护士;开展急症伤病机制、诊断、治疗、护理方面的研究,提高急诊服务质量,研究急诊质量监控。

3.1.2 急诊管理流程

1. 一般急诊诊疗程序

一般急诊诊疗程序如下:

- (1)挂急诊号的一般急诊患者,由分诊护士按内、外科分诊就诊。
- (2)如患者多,候诊时间较长,应及时报告科主任并安排增加急诊医师。
- (3)一般急诊诊疗结束后,值班医师完成急诊病历并记载注意事项,该病历交由患者保管,并在就诊登记本上记录。
- (4)凡不具备收入院条件,也不能院外观察的患者均应收留观。
- (5)急诊值班医师接诊到需要住院的患者时,完成必要的急诊检验检查。联系相关科室以便安排床位,及时开具住院证,优先安排住院。

图 3-1 所示是急诊科工作流程:

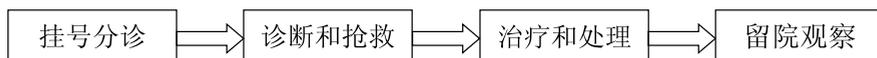


图 3-1 急诊科工作流程

2. 急危重症救治优先程序

为了保证急危重症患者抢救工作的及时、准确、有效,急诊中心开设并实施绿色生命安全通道,即“急救绿色通道”——对急危重症患者一律实行优先抢救、优先检查和优先住院原则。

(1)急危重症患者应先抢救,后挂号、交费、办理有关手续。

(2)急危重症患者抢救时,参加抢救人员必须以极端负责的态度,争分夺秒地抢救患者。

(3)一般抢救由急诊科医师、值班护士负责实施;如遇重大抢救应由科主任组织领导,或由医务处负责组织协调有关科室参加。

(4)在抢救过程中应有专人负责记录,要求准确、清晰、扼要、完整,并注明执行时间,抢救结束后及时整理、归档。

(5)预检护士发现急危重症患者应及时安排进抢救室,并立即通知医师进行抢救。

(6)抢救过程中医护人员密切配合,严格执行查对制度。

(7)经抢救病情稳定后,如须收入病房,由指定的医师、护士和担架员护送。

3. 急诊手术流程

急诊手术是指病情紧急,经医生评估后认为需要在最短的时间内实施的手术,否则患者就有生命危险。

病房急诊手术由病房医疗组组长或科主任决定,急诊室患者由当天值班最高级别医生决定,并遵照《手术分级管理及审批制度》执行。

(1)急诊手术流程

治疗医生发现患者需要急诊手术应立即请示医疗组组长或当天值班级别最高医生,必要时请示科主任。决定手术后,立即通知手术室、麻醉科。由急诊室尽快完成必要的术前检查、配血、术前准备。决定急诊手术后,主刀医师或第一助手应在急诊室详细向患者和/或家属说明病情、手术必要性、手术风险、替代治疗等情况,征得患者和/或家属签字同意。如患者因特殊原因(如昏迷)又无家属在身边,应报医务科或总值班审批。由手术医师、急诊科护士共同护送患者进入手术室。具体流程如图 3-2 所示。

(2)手术室急诊手术安排

为了保证急诊手术的及时安排,必须遵循以下原则:保留一间手术室为急诊手术专用,择期手术不得占用。同时有两台以上急诊手术,对于危及生命的

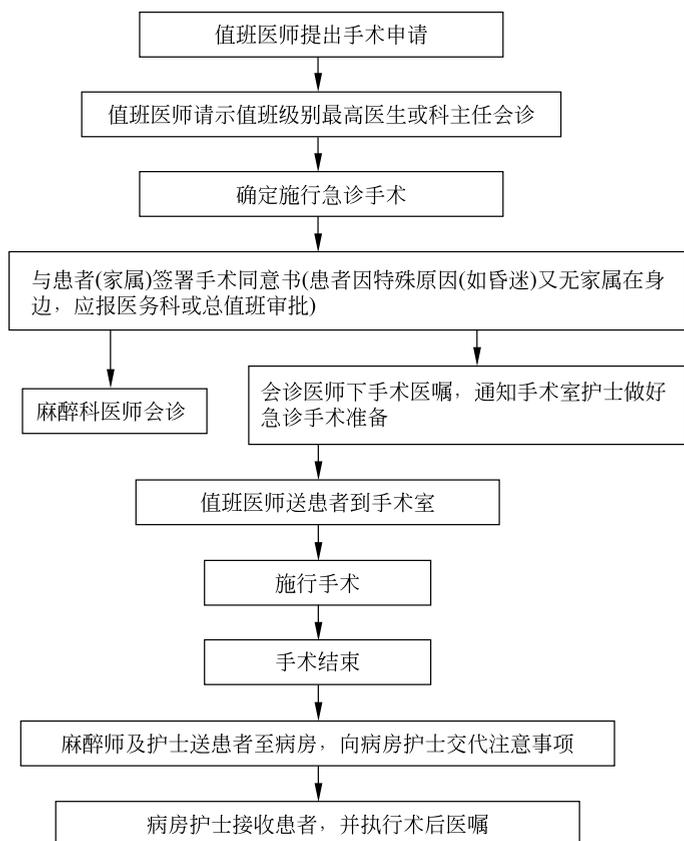


图 3-2 急诊手术流程

急诊手术, 手术室应立即以最短的时间安排接台, 由手术室护士长全权负责调配安排。非危及生命的急诊手术, 手术室根据情况安排接台, 原则上由手术室所属科室接台, 患者等待手术时间不得超过 2h, 急诊患者所属科室应在手术室安排手术台后半小时内将患者送至手术室。

3.1.3 急诊管理关键问题

急诊中心主要提供紧急医疗服务, 因为面对的紧急情况比较多, 所以经常出现拥堵和混乱的场景。医院在增加急诊医疗资源和优化资源配置的同时, 还必须对管理方法和手段进行优化, 以提高有限急诊医疗资源的效能, 提高患者满意度。

急诊服务管理中主要涉及的问题包括两个方面: 急诊患者调度问题和急诊手术调度问题。

1. 急诊患者调度问题

我国卫生部(2018年改为国家卫生健康委员会)在2012年9月发布的《医院急诊科规范化流程》中提出了急诊患者四级分诊标准:一级为濒危患者,需马上送至抢救室抢救;二级为危重患者,其病情有可能在短时间内进展至一级,或可能严重致残,应尽快处置及治疗,救治响应时间要求不超过10min;三级是急症患者,在短时间内没有危及生命或严重致残的征象,应在30min内安排患者就诊;四级是非急症患者,指患者目前没有急性发病症状,没有或仅有轻微不适,应在2h内安排患者就诊。到急诊中心的部分患者虽然症状较轻,但可能存在潜在的生命危险,及时安排诊治能大大降低患者死亡率和致残率。因此,如何通过急诊患者调度,实现急慢分诊、分级救治对于保障患者的生命安全、提高急诊服务质量至关重要。

急诊患者就诊时直接排队挂号,初诊、复诊患者(检查后再次回到诊室就诊的患者)就诊顺序不明确,拥堵造成患者等待时间长,所以首先要解决初诊患者和复诊患者的就诊顺序调度问题,一方面改善诊室就诊无序的现状,另一方面减少患者等待时间及急诊滞留人数。周鑫等^[1]利用诊室资源再分配提高紧急患者响应度。邓业雯等^[2]研究了初诊、复诊关键绩效指标、给出最优就诊顺序建议。Wen等^[3]采用马尔可夫决策过程(MDP)模型研究半紧急患者的实时调度问题,目标是最小化患者等待时间。

2. 急诊手术调度问题

手术室调度同时存在择期手术(本书第5章将着重讨论择期手术调度问题)和急诊手术,需要统筹分配资源。由于急诊手术的随机性和突然性,管理者不能在它们到达之前将急诊手术纳入计划,增加了医疗系统调度的不确定性和难度。因此分析急诊手术的特点,重新优化手术排程和资源,对手术室应急管理有实践指导意义。

增加急诊手术,使得原有已安排好的择期手术不再是最优方案,如何统筹两种不同手术(择期与急诊)的排程和资源分配问题,既保证急诊病人能优先安排手术,又减少择期手术的调整,使资源能合理利用,是统筹优化的关键。

童海星等^[4]在常规手术室排程优化问题基础上进一步考虑急诊手术特点,同时协调优化择期手术和急诊手术,以各类病人等待时间最短、关键资源利用率最高和同质资源利用率均衡为多目标决策,构建优化调度数学模型。王冰等^[5]研究了有急诊手术的单手术台调度问题,将问题看作三台机器的无等待排列流车间调度问题,为了减少随机到达的急诊手术对已有择期手术计划进度的

影响,同时提高效率和整个计划系统的稳定性,提出了一种预测反应式调度方法。刘清昭^[6]提出了全周期应急手术室应急管理思路,研究了针对突发事件下医院手术室应急管理及优化调度。

以下分别介绍这两个问题的最新研究成果及应用。

3.2 急诊患者调度问题

3.2.1 相关研究概述

急诊患者调度主要是确定急诊医生的下一个就诊的患者类别。Saghafian等^[7]按照患者就诊后是否需要入院治疗,将患者分流后采用不同的收治策略,并综合运用解析模型与仿真模型对分流是否能改善急诊性能做了研究。Saghafian等^[8]综合急诊患者的病情紧急程度与诊疗所需资源的复杂程度,将患者分为四类,建立了MDP模型,并求解分析得出最优的诊疗顺序。进一步地,Huang等^[9]同时考虑初诊、复诊患者不同的紧急程度及初诊患者响应时间要求,建立了传统的交通拥堵模型,分析得到了初诊和复诊患者进入诊室顺序的最优控制策略。Wen等^[3]考虑到患者在系统内接受服务和等待服务产生的成本,提出了一种用于实时调度的MDP模型,通过阈值策略安排初诊、复诊患者的就诊顺序。基于MDP特征分析,提出启发式策略和动态规划策略(approximate dynamic programming, ADP),并通过算例和实际应用说明其有效性。周鑫等^[1]利用Anylogic软件构建某三级甲等医院急诊部门仿真模型,使用FCFS、EDT、WEDT、WEDTL、PF等多种队列优先规则,提出通过二次分配医疗资源,为紧急患者设置“绿色通道”这一诊室资源再分配策略提高急诊对紧急患者的响应。邓业雯等^[2]采用Arena建立仿真模型,对初诊患者和复诊患者就诊顺序进行仿真研究,仿真分析得出等待时间、滞留时间、系统人数等几个关键绩效,同时根据这几个关键绩效给出最优的就诊顺序调度策略。

以下主要介绍基于仿真的方法和MDP模型。

3.2.2 问题分析与描述

患者进入急诊科后就诊流程如图3-1所示。急诊患者的就诊流程一般分为预检分诊、诊室初诊、检查、诊室复诊、患者分流五步。根据就诊流程,急诊患者分为初诊患者、检查患者与复诊患者。由于急诊患者的检查一般安排在当天,一般讨论的复诊患者是指当天内拿到检查报告返回诊室的患者。患者到达急诊中心后,先在预检台进行预检和分诊,挂号后进入候诊区等待进入诊室。初诊后的患者有一部分需要进行复诊,这部分患者先进行各项检查;如拍X光片、

B超、各种检验化验。需要接受检查的患者,进入不同的检查项目进行排队等候。检查结束后返回候诊区等待进入诊室复诊,不需要进行复诊和已经结束复诊的患者可离开急诊中心。

邓业雯等^[2]以上海某三甲综合医院 2014—2015 年急诊内科的数据为基础,将一天按照小时分为 24 个时间段,统计得出各个时间段内的平均到达率,然后,用 SPSS 统计软件对每个时间段的患者到达数量进行拟合,拟合结果表明,初诊患者的到达近似服从泊松分布。就诊人数呈现早晚高峰的现象,早高峰出现在 7:00—8:00,晚高峰出现在 19:00—21:00。用泊松分布拟合复诊到达,拟合结果显示,泊松分布能够较好地拟合大部分情况,但是在 8:00—9:00、9:00—10:00 这两个阶段误差较大。主要原因是:高峰期患者在检查、缴费、等待检查报告等环节滞留时间较长。将初诊患者的等待时间近似处理为:等待时间=接诊时间-挂号时间-5min,统计发现 7:00—8:00 的等待时间较长,这与该阶段较高的到达率有关系。平均等待时间是指患者从进入急诊科(系统)到见到医生之前的等待时间平均值;调度策略的变化对初诊患者的等待时间影响不是很显著,而对于复诊患者而言,其等待时间会显著受到调度策略影响,即按照初诊和复诊患者 1:1 的调度策略进行叫号,可以使平均等待时间、系统时间(患者从进入急诊科(系统)到离开的时间)以及系统人数(在急诊科(系统)中等待和就诊人数)相对最少。

患者调度的规则有很多种,各种规则效果如何?“绿色通道”的使用是不是能够提高资源的效率?不同类型的患者应该如何调度?以下将就这些问题进行讨论。

3.2.3 基于仿真的调度方法^[1]

1. 模型建立

急诊系统中患者队列优先规则主要集中在三个方面:初诊队列优先规则、初复诊患者放行规则以及检查队列规则。一般医院排队叫号系统采用的初诊队列优先规则是先到先服务(first come first served,FCFS)策略,但该策略无法实现患者分级。最早交付日期优先规则(earliest due time first,EDT)及患者等级优先规则(priority first,PF)可以实现患者分级。

EDT 规则优先安排等待时间即将超过目标时间的患者,以提高其初诊前等待时间(waiting time before treatment,WTT)达标率,优先级只取决于患者等待时间与 WTT 的差值,不能很好地体现出患者分级的作用。

周鑫等^[1]提出两种用于对比的队列规则,分别称为加权目标等待时间优先

规则(weighted earliest due time first, WEDT)及考虑等级的加权目标等待时间优先规则(weighted earliest due time with patient level first, WEDTL)。

设急诊排队系统中的患者等级为 $L_i = i (i=1, 2, 3, 4)$, 其数字越小, 紧急程度越高。患者进入急诊系统的时间为 E , 当前时刻为 t , W_i 为患者的 WTT(单位为 h), 其中 $W_1=0, W_2=1/6, W_3=1/2, W_4=2, k$ 为常数, 取值大于 4。则患者的初诊队列内 WEDT 和 WEDTL 优先级为

$$P_{\text{WEDT}} = \frac{t - E}{W_i} \quad (3.2.1)$$

$$P_{\text{WEDTL}} = \frac{t - E}{W_i} (k - L_i) \quad (3.2.2)$$

这两个策略分别对患者 WTT 和等级进行了加权; WEDT 策略针对患者的 WTT 对其排序进行了加权; WEDTL 策略则增加了对患者等级的加权。该研究提出对现有医疗资源进行二次分配, 设置“绿色通道”诊室, 即将现有诊室分成普通诊室与“绿色通道”诊室, 两类诊室使用不同的接诊逻辑。“绿色通道”诊室除了服务初诊的二级患者外, 在初诊队列中没有紧急患者时, 还需收治队列中的其他等级患者以分担普通诊室的就诊压力。初、复诊患者放行比例为 1:1。针对在检查队列中的患者, 则采用 FCFS 和 PF 规则。

2. 算例

基于 Anylogic 仿真模型进行仿真, 仿真中涉及的队列规则和优化方案包括: FCFS、EDT、WEDT、WEDTL、PF 等初诊队列规则, FCFS 和 PF 等检查队列规则, 以及“绿色通道”(用“G”表示)诊室资源分配方案。设置 Anylogic 仿真时间单位为 1h, 单次仿真时间长度为 80 000h, 仿真人次为 100 万次, 使用的策略组合共计 18 种, 其中 8 种方案使用了“绿色通道”。如表 3-1 所示。

仿真中, 24 个时段的患者到达率服从泊松分布, 其中 1~4 级患者的占比分别为 1.66%、1.82%、23.35%、73.17%, 1 级患者与 44.01% 的 2 级患者直接进入抢救室。患者在诊室内平均就诊时长为 15min, 就诊时长服从指数分布。患者中需要复诊的比例为 59.15%。复诊前患者需要接受各项检查, 其中 92.69% 的患者进行血液检查、22.12% 的患者接受超声检查、接受 CT 和 X 光检查的比例分别为 55.33% 和 28.91%。

仿真结果表明: 在初诊队列中应用 FCFS 策略会使大多数紧急的 2、3 级患者不能在要求时间内就诊。EDT、WEDT 和 WEDTL 三种初诊队列策略在改善紧急患者的 WTT 达标率上很有效, 其中 WEDTL 策略效果最佳, 2、3 级患者的 WTT 达标率分别达到了 68% 和 82% 以上, 且 4 级患者的 WTT 达标率没有明显下降。在初诊队列中使用 PF 规则能够显著提高 2、3 级患者的 WTT 达标

率,但在 PF 规则下 4 级患者的优先级最低,将大幅降低其 WTT 达标率。在实际急诊管理中 PF 规则不适用,4 级患者虽然病情不紧急,频繁的插队将会带来投诉、患者满意度下降等问题。现有参数设置下,在初诊队列中使用 PF 规则,2、3 级患者的 WTT 达标率已经到达上限。而 PF 策略与“绿色通道”策略结合后,紧急患者的 WTT 接诊率没有统计学意义上的提升,反而 4 级患者的指标有明显下降。

表 3-1 仿真使用的策略组合表^[1]

策略命名	初诊队列规则	检查队列规则	使用“绿色通道”方案
F-F	FCFS	FCFS	否
F-P	FCFS	PF	否
EDT-F	EDT	FCFS	否
EDT-P	EDT	PF	否
W-F	WEDT	FCFS	否
W-P	WEDT	PF	否
WL-F	WEDTL	FCFS	否
WL-P	WEDTL	PF	否
P-F	PF	FCFS	否
P-P	PF	PF	否
EDT-F-G	EDT	FCFS	是
EDT-P-G	EDT	PF	是
W-F-G	WEDT	FCFS	是
W-P-G	WEDT	PF	是
WL-F-G	WEDTL	FCFS	是
WL-P-G	WEDTL	PF	是
P-F-G	PF	FCFS	是
P-P-G	PF	PF	是

在拥挤程度方面,最拥挤的方案是 WL-P-G 方案,这种策略下系统内平均人数达到了 43.93;W-F 方案拥挤程度最低,系统内平均人数为 37.27。在应用了“绿色通道”方案后,急诊系统拥挤程度加剧,这是由于“绿色通道”诊室的设置使得更多的非紧急患者在初诊队列中等待。实验结果表明检查队列使用的两种规则无明显差异,原因是检查的等待时间占患者总等待时间的比例很小。在负荷较高的急诊系统内,提高对紧急患者的响应,通常伴随着对非紧急患者的响应变差,在急诊系统当中为紧急患者让渡一些资源和时间是相对合理的。

使用不同策略对诊室资源再分配,不仅对患者的初诊 WTT 平均达标率有提升作用,还对由于一天中患者到达率波动带来的系统压力有所平衡。开设“绿色通道”诊室缓解了由于到达率波动带来的高峰就诊压力,提高急诊高峰时段的响应性,减少因等待发生的投诉和冲突。

3.2.4 基于 MDP 的调度方法^[3]

1. 模型建立

考虑急诊中心半紧急分类患者(triage patients, TP)和流程中检验患者(inprocess patients for examination, IPE)的动态调度。对于一个医生,将其一天工作时间分成若干相等的段,用 $t(t=1, 2, \dots, n)$ 表示。不同级别患者到达挂号等待就诊,就诊后分流,一部分离开系统,另一部分进入检验队伍排队。问题是要确定 TP 队伍和 IPE 队伍里下一个接受服务的患者是谁。

在 t 时段,不同级别急诊患者进入 TP 队列, $a_{-i,t}$ 表示 WTT 为 i 的 TP 患者数目,令 $a_t = \{a_{0t}, a_{-1,t}, \dots, a_{-l,t}\}$, $x = (x_{t-1,t}, x_{t-2,t}, \dots, x_{-l,t})$ 表示推迟了 i 时间的 TP 患者数, y_t 表示流程中等待诊疗患者(inprocess patients for treatment, IPT)队列长度, z_t 是 IPE 队列长度。调度决定进入服务的下一个患者是哪位。如果是 TP 患者,则他在时段 t 末以概率 p_1 结束问诊,或以概率 $1-p_1$ 回到 TP 的等待队列。如果诊疗结束,TP 患者以概率 p_d 离开系统,或以概率 $1-p_d$ 进入检查队列。如果一个 IPT 患者被选择,则他在时段 t 末以概率 p_2 结束问诊,或以概率 $1-p_2$ 回到 IPT 的等待队列。对于每个 IPE 患者,在时段 t 末他会以 p_3 概率进入 IPT 等待队列,或在下一阶段以 $1-p_3$ 概率继续他的检查。其转移概率如图 3-3 所示。

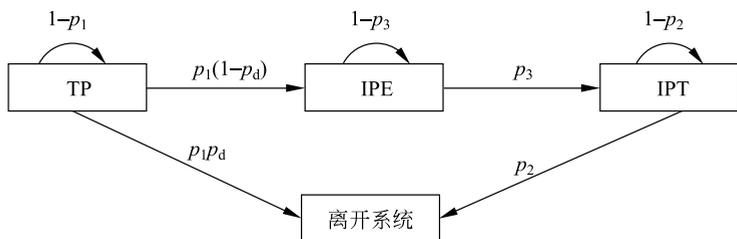


图 3-3 半紧急患者服务过程转移概率^[3]

目标是要使得总期望成本最小化。 β 表示 IPT 的边际等待成本, α_i 表示 TP_i 的边际等待成本,且有 $0 = \alpha_{-l} \leq \alpha_{-l-1} \leq \dots \leq \alpha_0 \leq \beta \leq \alpha_1 \leq \dots \leq \alpha_{t-1}$ 。不失一般性,令 $\beta=1$ 。这个成本结构对于 TP 患者最优调度策略为最早到期策略(earliest-due-date first, EDD),为了简化模型,在第一个问诊结束离开前,对 TP

队伍设置等待成本。WTT 在模型中是最大允许的等待时间和平均服务时间之和。将实时动态调度问题建立 MDP 模型,时段 t 系统的状态用向量 $s_t = \{x_t, y_t, z_t\}, t \geq 1$ 表示。在每个阶段要确定下一个服务的患者。用 $n_t \in \{0, 1\}$ 表示 TP 患者在 t 时段是否得到服务。在 t 时段获得服务的 IPT 数用 $(1 - n_t, y_t)$ 表示。状态转移函数为^[3]

$$x_{-I,t+1} = a_{-I,t+1} \quad (3.2.3)$$

$$x_{i+1,t+1} = \begin{cases} x_{it} + a_{i+1,t+1}, & \text{如果 } i \neq i^* \text{ 且 } -I \leq i < 0 \\ x_{it}, & \text{如果 } i \neq i^* \text{ 且 } 0 \leq i < t \\ (x_{t^*t} - n_t \eta_1)^+ + a_{i^*+1,t+1}, & \text{如果 } i = i^* \text{ 且 } -I \leq i^* < 0 \\ x_{t^*t} - n_t \eta_1, & \text{如果 } i = i^* \text{ 且 } 0 \leq i < t \end{cases} \quad (3.2.4)$$

$$y_{t+1} = y_t - \min(1 - n_t, y_t) \eta_2 + r_t \quad (3.2.5)$$

$$z_{t+1} = z_t - r_t + \eta_1 \delta n_t \quad (3.2.6)$$

此处, $i^* = \max\{i: x_{it} > 0\}$ 。 η_i 服从参数为 p_i 的伯努利分布,代表 TP 队列 $i=1$ 和 IPT 队列 $i=2$ 问诊结束概率分布。 r_t 服从参数为 z_t 和 p_3 的二项分布,代表完成初诊回来进行复诊的患者数, δ 服从参数为 p_d 的伯努利分布,代表患者是否需要检查。

阶段成本在时段 t 初始时开始计入,表示为

$$g(s_t) = \sum_{-I \leq i < t} \alpha_i x_{it} + y_t \quad (3.2.7)$$

目标函数是要最小化总成本,即当 $T' > T$ (T 为总时间段数)和 $S_{T'} = 0$ 时,求 $\min E \left[\sum_{t=1}^{T'} g(s_t) \right]$ 。建立 MDP 模型,最优等式为

$$V_t(s_t) = g(s_t) + \min E[V_{t+1}(s_{t+1})] \quad (3.2.8)$$

Wen 等^[3]对以上模型特性进行分析,并设计 ADP 方法求解以上模型。同时,对比医院目前使用的调度策略和几个启发式策略。

2. 算例

使用某三级甲等医院 2016 年和 2017 年急诊中心患者记录,急诊中心医生 24h 轮班工作,因晚上到达率较低,故仅选取了每天 8:00—18:00 的记录。6min 为一个时间段,一天中有 100 个时间段,即 $T = 100$ 。假设一个医生在一天中所有时段均可以问诊,半紧急患者的到达率是不稳定的。一个新到达的 TP 患者可为 3 个等级中的某个,故令 $p_u = 0.3$ 。TP 和 IPT 的平均服务时间分

别为 9min 和 15min。因此, TP 和 IPT 每个时段完成服务的概率是: $\frac{2}{3}$ 和 $\frac{2}{5}$, 即 $p_1 = \frac{2}{3}$ 和 $p_2 = \frac{2}{5}$ 。在 TP 患者中, 大约 40% 在问诊后去进行检验检查, 大约 63min 后返回进行复诊。因此设置 $p_d = 0.4$, $p_3 = 0.095$ 。参照医院历史数据, 等级 3 和等级 4 患者的 WTT 分别为 30min 和 2h。将平均服务时间(为 9min)加上 WTT, 分别为 39min 和 129min。

设单位 IPT 等待成本为 1, $c_i = \gamma (i > 0)$, 一个 TP 的 $\gamma = 6$ 。这里的 TP 成本结构并未考虑其服务时间超出或低于 WTT 的惩罚。实际数据应用到该 MDP 模型很难求解, 因此采用启发式策略以对比其效能。设计一个较小规模的问题, WTT 针对等级 3 的患者, 设置为 4 阶段(24min), 等级 4 患者设置为 6 阶段(36min)。同时, TP 到达率减少为原来的 $\frac{2}{3}$ 。

对于这个小规模问题, 分析了最优控制策略和其他策略的性能对比, 并进行了敏感度分析。

表 3-2 是小规模问题第 85 时段的最优控制。

表 3-2 85 时段最优控制^[3]

TP ₂ /TP ₄	0	1	2
0	IPT	IPT	IPT
1	IPT	IPT/TP	TP
2	TP	TP	TP

表中 TP₂ 表示等级为 2 的 TP 患者, TP₄ 表示等级为 4 的 TP 患者。表中显示两种类别患者在 0~2 之间变动时的调度策略, 如两类都是 0 时, IPT 进入队列, 两类都是 1 时, IPT 和 TP 交替进入队列。研究同时对比了不同策略下总成本(TC)、总成本 95% 置信区间的半宽度(CI)、TP 等待成本(TPcost)和 IPT 成本(IPTcost)、TP 平均等待时间(TPLoS)、IPT 平均等待时间(IPTLoS)、所有患者的平均等待时间(AvgLoS)、TP 患者平均延迟率(Perc_Tardy)、插队 IPT 概率(P-IPT)、插队 TP 的概率(P-TP)、超时、运行时间、最优性差距 $Gap = (TC_{Heuristic}/TC_{Opt}-1) \times 100\%$, 数据如表 3-3 所示。

最优性策略中, 其总成本 $TC = 165.15$, 是最低的, 而平均延迟率 Perc Tardy 为 5.03%, 说明它能以近 95% 的概率保证 TP 在 WTT 时间内得到服务。

改变参数进行敏感度分析, 发现随着等级 3 的患者概率(p_u)、TP 患者到达率(p_n)、IPE 服务率(p_3)和 TP 单位延迟成本(γ)增加, 总成本增加; 但随着 TP 服务率(p_1)、IPT 服务率(p_2)和 TP 离开率(p_d)增加而总成本减少。

表 3-3 基础案例的不同策略性能比较^[3]

策略	Opt	TIT	TP-first	Slack ($k^* = 2$)	ETT ($\omega^* = 0.5$)	Look2-sc	Look2-ETT
TC/min	165.15	461.98	196.25	167.53	166.29	173.36	165.15
CI/min	0.35	1.29	0.38	0.35	0.35	0.37	0.35
TPcost/min	19.91	389.67	24.09	21.83	25.62	39.61	19.91
IPTcost/min	145.2	72.3	172.2	145.7	140.7	19.9	172.2
TPLoS/min	22.26	35.58	16.44	22.2	23.28	24.6	22.26
IPTLoS/min	77.64	66.18	80.94	77.7	77.1	76.08	77.64
AvgLoS/min	55.5	53.94	55.14	55.5	55.56	55.5	55.5
Perc Tardy/%	5.03	22.50	4.11	4.85	6.23	8.73	5.03
P-IPT	0.30	0.00	0.00	0.26	0.28	0.23	0.30
P-TP	0.02	0.00	0.00	0.03	0.03	0.03	0.02
超时/min	123.8	136.4	118.3	122.9	124.3	125.2	123.8
运行时间/s	1346	14	14	105	96	551	558
Gap/%	—	179.73	18.83	1.44	0.69	4.97	0.00

注:表中各种策略描述请参阅文献[3]。

为了验证启发式策略的可行性,从医院搜集实际运营数据进行验证。服务时间采用了几何分布和指数分布。在几何服务时间情形下,最小成本的策略是ADP方法,即Look2-ETT策略。在随机指数服务时间情形下,ETT策略和Slack策略总成本最小。

3.2.5 管理启示

本部分介绍了两种急诊患者调度方法,分别是基于仿真的方法和基于MDP的方法。两种方法都是针对当前急诊患者分级情况下讨论调度不同级别患者的策略。在管理上可以借鉴以下方法:

(1)急诊患者分级方法,根据已有研究,有两种分级方式:第一种分初诊和复诊两种类型;第二种按照紧急程度分四级。第一种分类方式是以第二种为基础的,相对简单,在进行患者调度时相对容易;第二种分类方式需要结合医学知识进行判定。

(2)不同调度策略的采用,在已有研究里提出了多种调度策略,不同调度策略对患者等待时间和总成本影响不同,复杂的如启发式策略等需要有相应计算

机软件和计算能力支持。

(3)绿色通道的使用,为了保证特别紧急患者能够及时得到救助,有学者提出绿色通道策略,但在应用了“绿色通道”方案后,急诊系统会变得更加拥挤,这是由于“绿色通道”诊室的设置使得更多的非紧急患者在初诊队列中等待,所以,如果使用绿色通道,应该及时进行调度,适时调整非紧急患者进入绿色通道以降低拥挤状态。

3.3 急诊手术调度问题

3.3.1 相关研究概述

在手术室的规划和调度问题中,择期手术和急诊手术是两种不同方式的手术^[10]。择期手术由于其完整的、已知的信息可以定期计划和安排。现有的大多数手术调度研究工作主要关注择期手术的操作^[11-13],择期手术的调度时间表被视为外科医生、护士和患者为即将到来的手术做准备的依据。急诊外科病例在医院经常出现,急诊手术通常以挽救生命为目的,需要尽快处理。但是,管理者不能在它们到达之前将其纳入调度计划。急诊手术的随机到来增加了医疗系统的不确定性^[14]。

前期关于急诊手术调度研究有许多,一部分是仅考虑急诊手术本身调度问题,如 Van Essen 等^[15]提出在择期手术计划中插入急诊手术的方法,以减少急诊手术等待时间,他们提供了一种方法,将急诊手术的“插队时间”尽可能均匀地分布在一天中。Stuart 和 Kozan^[16]研究了手术室日常运行的中断管理和重新安排问题。他们对在线环境中的择期和非择期患者进行排序,以最大限度地增加预计按时完成的手术的加权数量。Lambrechts 等^[17]采用基于时间松弛的技术,在急诊手术到达之前,以计划的时间表来适应即将到来的急诊手术。还有些学者将机器调度理论运用到急诊手术调度中,Fei 等^[13]将日常手术调度问题视为一个两阶段的混合 flowshop 问题,并采用混合遗传算法进行求解。Zhong 等^[18]将手术调度问题视为并行机调度问题,提出了一种两阶段方法。因为急诊患者随机到达手术室的时间早于计划的执行时间,王冰等^[5]提出了一种新的预测反应调度方法,以适应急诊患者的到来,同时采用效率和稳定性相结合的目标函数,在生成择期手术离线计划时间表时,考虑了随机插入急诊患者的影响,将部分重调度方法应用于急诊患者到达后择期手术的反应调度阶段。童海星等^[4]在常规手术室排程优化问题基础上进一步考虑突发事件特点,同时协调优化择期手术和急诊手术,以各类患者等待时间最短、关键资源利用率最高和同质资源利用率均衡为多目标决策,构建优化调度数学模型。

以下将着重介绍预测反应调度方法和突发事件下手术室应急调度模型。

3.3.2 问题分析与描述

通常情况下,手术室执行计划的时间表,当急诊患者到达时,一种立即处理急诊患者的方法是让急诊患者进入计划的择期患者队列,这样将使得原有已安排好的择期手术计划不再是最优。如何统筹两种不同手术(择期与急诊)的调度和资源分配问题,既保证急诊病人能优先安排手术,又减少择期手术的调整及提高资源合理的利用率,是统筹优化的关键。

急诊手术的到来通常是不可预测的,因其突发性和紧迫性,且急诊患者规模不同,手术量也会不同程度地增加,这必将冲击原来计划的择期手术排程。所以应急调度问题将着重研究激增的紧急手术对原有择期手术带来的冲击。常规情况下手术排程优化研究,大都是以总时长最短、医护人员的满意度最高、成本最低等作为优化目标,但是突发事件下的应急调度问题,与常规排程调度问题的优化目标有着很大的不同。

在手术过程中,典型的患者经过三个手术阶段:术前阶段、术中阶段和术后阶段,分别在术前监护室(preoperative holding unit, PHU)、手术室(operation room, OR)和麻醉后监护室(post anesthesia care unit, PACU)进行,如图 3-4 所示。每一个手术在三个手术阶段都不能被打断,一旦手术开始,必须连续进行三个阶段,其间没有等待时间,称为无等待约束。三个手术阶段的手术时间可能不同,每个阶段所需的准备时间计入相应阶段的手术时间。手术室的一个可行的时间表是在三个手术阶段之间分配手术时间,满足无等待约束,目标是最小化完工时间。

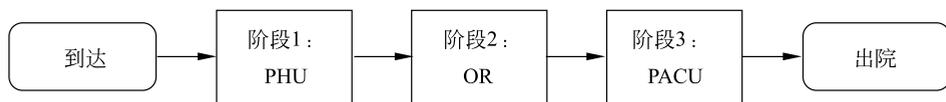


图 3-4 外科手术患者流向^[5]

可以通过历史数据统计计算急诊手术的发生概率,在此基础上进行手术调度方法研究。

当考虑急诊手术的突发性时,应考虑以下两个关键概念^[4]:

(1)患者复合等待时间:复合等待时间综合考虑不同手术的紧急度和患者等待时间。突发事件带来的一些患者由于伤情紧迫,需要即刻手术安排,延迟手术可能意味着病情的急剧恶化甚至有生命危险,所以将不同病情程度(即紧急度)和等待时间综合考虑是急诊手术调度的紧要目标;

(2)关键资源利用率:在常规情况下,资源的优化调度目标是均衡,这样避

免资源集中,在突发事件下,激增的手术需求一般会集中在某个科室类型上,此科室的资源视为关键资源,需要充分利用,且需要整合其他资源以完成手术,这样才能救治更多的急诊患者。

3.3.3 预测反应调度方法^[5]

1. 模型建立

1) 择期手术调度计划

考虑 n 名择期患者安排在一个手术室,所有患者手术完成时间是最后一台手术完成时间,时间表可以由所有手术的排列表表示。设 $s = (s_1, s_2, \dots, s_n)$ 表示 n 个手术的排列,其中 $s_i (i=1, 2, \dots, n)$ 表示排在第 i 个位置的手术。设 S 是 n 个手术的所有可能排列的集合。设 $b(s_i, j)$ 为手术排列中 s_i 第 j 个手术阶段的手术开始时间,其中 $j=1, 2, 3$ 。设 $p(s_i, j)$ 为第 j 阶段手术 s_i 的持续时间, $f(s_i, j)$ 为第 j 阶段手术 s_i 的完成时间。

由于要满足无等待约束,当为排列 $s = (s_1, s_2, \dots, s_n)$ 生成手术的时间安排表时,对于任何 $j=2$ 或 3 ,必须满足 $b(s_i, j) = f(s_i, j)$ 。无等待约束的要求可能会导致两种后果:第一,为了保证无等待约束,三个手术阶段之间可能存在空闲时间,如图 3-5 所示;第二,后两个($j=2, 3$)阶段的手术时间分配将由第一个($j=1$)阶段的手术时间分配决定。因此,所有手术的具体时间分配实际上是由所有手术的排列以及第一个手术阶段手术之间存在的空闲时间量决定的。

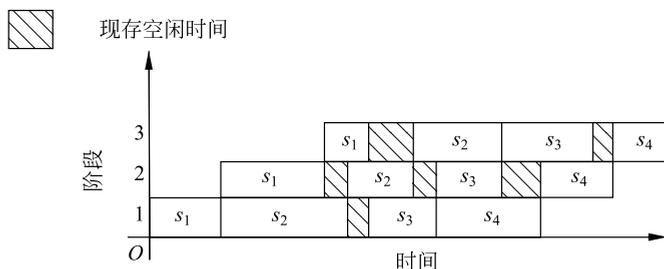


图 3-5 无等待约束下三个阶段手术之现存闲置时间^[5]

三个手术阶段中第一个手术的开始时间可确定如下:

$$b(s_1, j) = \begin{cases} 0, & j = 1 \\ \sum_{h=2}^j p(s_1, h-1), & j = 2, 3 \end{cases} \quad (3.3.1)$$

设 $I(s_i, j)$ 为第 j 个操作阶段($j=1, 2, 3$)的手术 I 与下一个手术 $I+1$ 之间的空闲时间量, $I(s_i, 1)$ 为第一个手术阶段的手术 s_i 与下一个手术 s_{i+1} 之间

的空闲时间量。因此, $I(s_i, 1)$ 可由以下公式计算:

$$I(s_i, 1) = \max \left\{ 0, \max_{k=2,3} \left[\sum_{j=2}^k p(s_i, j) - \sum_{j=1}^{k-1} p(s_{i+1}, j) \right] \right\}, \quad i = 1, 2, \dots, n-1 \quad (3.3.2)$$

第一个手术阶段的手术开始时间可通过以下公式确定:

$$b(s_i, 1) = \sum_{l=1}^{i-1} [p(s_l, 1) + I(s_l, 1)], \quad i = 2, 3, \dots, n \quad (3.3.3)$$

因此,所有手术的完成时间为

$$f(s_i, j) = \begin{cases} \sum_{h=1}^j p(s_i, h), & i = 1; j = 1, 2, 3 \\ b(s_i, 1) + \sum_{h=1}^j p(s_i, h), & i = 2, 3, \dots, n; j = 1, 2, 3 \end{cases} \quad (3.3.4)$$

其中:

$$f(s_n, 3) = \sum_{l=1}^{n-1} [p(s_l, 1) + I(s_l, 1)] + \sum_{h=1}^3 p(s_i, h) \quad (3.3.5)$$

时间表 s 的最大完工时间正好是最后一台手术 s_n 的完成时间。用 $M(s)$ 表示最大完工时间:

$$M(s) = f(s_n, 3) \quad (3.3.6)$$

$M(s)$ 可以作为手术调度时间安排表的效率参数,优化单一手术组调度效率,可以生成择期手术的初始时间表。一般情况下,在不考虑任何不确定性的情况下生成的初始计划可以作为外科医生、护士和患者为即将到来的手术提前准备的择期手术计划。

2) 急诊手术调度计划

进一步考虑单一手术室受随机急诊手术约束的计划制订。通常情况下,手术室执行择期计划时间表,直到急诊患者到达,通常让急诊患者进入计划的择期患者队列。

假设一天中最多只有一个急诊手术,三个手术阶段工作持续进行。通过历史数据统计分析获得急诊手术的发生概率。基于已知的急诊手术发生概率,研究预测性的反应式调度方法,生成预测时间表以替代前面生成的初始时间表。手术室执行相应的预测时间表,直到急诊患者到达重新调整受影响的择期手术。实际的手术调度方案为急诊手术前已经执行的部分与插入急诊手术后重新安排的解决方案合并。