Scheduling operating rooms in overcrowded Chinese hospitals using a sample average approxatimation method

guanlian xiao, wllem van jaarsveld, joris van de klundert

ExtendedAbstract

The economic and societal progress made in the People's Republic of China since 1978 have been enormous. GDP rose from 189 billion USD in 1980 to 9240 USD in 2013 (World Bank, 2015). Although health reforms have lagged behind especially in rural areas (China was ranked 144th among 190 countries by WHO on health equity in 2000), large scale health reforms have been initiated in the present millennium (Chen, 2009). Between 1995 and 2010, China's government budget on health grew tenfold (Liang & Langenbrunner, 2013). The economic growth and increased health insurance coverage have stimulated health service demand. Many hospitals however, especially in rural areas, experience difficulties to meet the growing demand, and demand for more complex health services, due to limitations in resources, in particular the availability of qualified and well trained human resources. The resulting limitations in capabilities cause many rural citizens to bypass local health service facilities to avoid misdiagnosis and under-qualified staff and facilities (Shi, 2014). Instead, they seek service in renowned city hospitals, which also need to cope with rapidly increasing urban populations (bypassing urban primary care facilities). As a result, especially the larger (level 3) hospitals in big cities face capacity problems and are severely overcrowded. In fact, the hospitals have become markedly overstretched and many patients have been unable to gain access to treatment. In this paper we investigate the resulting capacity management problems on the basis of a case study in the First People Hospital in Shanghai. The study focuses on the capacity management problems in the operating theatres.

While operating room capacity is already over demanded, it also suffers from physicians and other (human and non-human) resources being pulled away, mostly by outpatient departments. The operating theatres of large big city hospitals have therefore become particularly exposed to the high pressures of rising demands. At Shanghai's First People Hospital for instance, the actual average surgical workload exceeded the daily capacity of 9.5 hours by as much as 20 percent in 2013. As idle time is in the order of 17 percent, the actual opening hours of the operating theatres are almost 14 hours per day. In this paper we consider the thus arising overtime work, in relation to idle time. Our research aim will be to investigate effective improvement operating room scheduling methods, which take idle time and overtime work

into account. Moreover it considers one of the root causes of idle time: cancellation.

In China cancellation reported cancellation rates have typically been between10 and 15 percent and caused by hospital acquired infections, inappropriate preoperative preparation, poor scheduling, and unaffordability of out of pocket payments by patients (Jiang et al.,2011, Jie, et al. 2012,Li et al. 2011, Xiang et al. 2014, Zhang et al. 2014, Xu et al., 2009). Next to these executional reasons, several authors mention the lengths of schedules and workload as reasons for (end of day) cancellations. Such end of day cancellation may cause dissatisfaction and loss of health for patients scheduled for surgery. On the other hand, longer overtime hours have been reported to be associated with increased risks of complications and medical errors, as well as dissatisfaction among scarce staff.

In this paper, we optimize schedules taking patient and resource unavailability into account, as well as the costs of (late) overtime work. Patient and resource unavailability are modelled as independent stochastic processes. The surgery durations are modelled as random variables. We consider the stochastic scheduling problem of finding optimal schedules, where optimality is defined in terms of maximizing the (health) value created by operating patients minus the cost (and risks) of overtime work. The developed solution method for this scheduling problem, as further detailed out below, realistically includes the opportunity to revise scheduling decisions after the morning session has ended and before the afternoon session begins. The purpose will firstly be to see if and how optimization models which explicitly consider timely cancellation (at lunch time instead of at end of day), yield better solutions. This first purpose contributes to taking away one of the frequently cited sourced of cancellation: poor scheduling. A second purpose will be to explore the potential of alternative sources of cancellation which management decisions might (partially) eliminate. We explore how reductions in resource and patient inavailability can improve the combined objective function of creating more (health) value and reducing the 'costs' of overtime work. Our research therefore extends the work of **?**, who introduce 'patient rejection' based on actual realizations, in a simulation framework, and appear to be the first explicitly consider cancellation decisions in operating room scheduling.

We formulate the operating room scheduling problem as a general two stage stochastic program with *integer recourse*. The recourse concerns the opportunity to cancel patients originally scheduled for surgery in the afternoon, after learning the realizations of the duration

for the patients scheduled in the morning. (Such cancellation is considered to be 'on time' for managing alternatives for patients for which it is necessary, and to inform patients for which surgery is not urgent about the delay.)

Standard two-stage stochastic programs with linear or convex functions are often solved using the L-shaped method or Bender's decomposition (see for instannce (Birge, 1993) and (Van Slyke & Wets, 1969). More complex two stage stochastic program with integer recourse are covered by Laporte et al. (1993), or Angulo (2014). Following the approach of Laporte et al. (1993), we exploit the special relationship between first stage realization and optimal second stage integer solutions to solve our problem. Moreover, we use Jensen inequality (**?**) to accelerate the computation as proposed in Batun et al. (2011). We integrate sample average approximation (SAA) and the L-shaped method to solve the resulting two stage stochastic integer programs **?**.

As Stepaniak et al. (2012) showed that the lognormal distribution best fits empirical surgical procedure times data in Dutch hospitals, we have fit lognormal distributions to the empirical data from Shanghai First People Hospital, and verified the fit. Computationally, the lognormal distribution is cumbersone as a closed form expression for the distribution function of the sum of lognormal distributions is unknown. Our Stochastic Average Approximation model is adjusted to deal with lognormal surgery duration (as well as a bernoulli distributed patient inavailability).

We speed up the process through several analytical results which assist us to strenghten the formuatlions in the the form of optimality cuts. We show that for a given solution of the master problem, the expected number of cancellations is decreasing in the sum of the realized procedure trimes for the patients scheduled in the morning. This result enables us to find valid inequalities for the master problem, using Jensen's inequality. Moreover, we prove that patients whose distributions have a smaller mean and smaller variance are scheduled earlier.

The developed algorithm can solve the daily scheduling instances, and yields answers for the research questions stated above. The computational analysis shows that the overall improvement of the objective function as can be achieved from cancellation is between 3 and 9 percent for test instances which are derived from the empirical data. Moreover, our results indicate that the methods can deal relatively well with patient cancellations, as the objective function improvements attainable by reducing patient no show are very small. Interestingly,

improving the resource availability, can lead to considerable improvements of the objective functions. If the current expected unavailability of around 60 minutes per day (slightly moren than 10 percent of the available time) is reduced to zero, the objective function value increases by approximately 28 percent for the test instances. Depending on the cost of overtime work, the number of cancellations doesnt necessariliy follow the same trend as the objective functions considers them jointly with the value of the scheduled patients.