ARTICLE IN PRESS

Journal of Operations Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Worth the wait? How restaurant waiting time influences customer behavior and revenue

Jelle De Vries[a,b,*], Debjit Roy[b,c], René De Koster[b]

[a] VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[b] Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
[c] Indian Institute of Management, Vastrapur, Ahmedabad, Gujarat 380015, India

ARTICLE INFO

Keywords:
Waiting time
Customer behavior
Transaction data
Revenue
Simulation

ABSTRACT

In many service industries, customers have to wait for service. When customers have a choice, this waiting may influence their service experience, sojourn time, and ultimately spending, reneging, and return behavior. Not much is known however, about the system-wide impact of waiting on customer behavior and resulting revenue. In this paper, we empirically investigate this by analyzing data obtained from 94,404 customers visiting a popular Indian restaurant during a 12 month period. The results show that a longer waiting time relates to reneging behavior, a longer time until a customer returns, and a shorter dining duration. To find out the impact of the consequences of waiting time, we use the empirical findings and data collected in a simulation experiment. This experiment shows that, without waiting, the total revenue generated by the restaurant would increase by nearly 15% compared to the current situation. Stimulating customers to reserve could enable restaurants to reap part of this benefit. Furthermore, the results of simulation experiments suggest that, within the boundaries of the current capacity, revenue could be increased by a maximum of 7.5% if more flexible rules were used to allocate customers to tables. Alternatively, by increasing the existing seating capacity by 20%, revenue could be boosted by 7.7% without the need to attract additional customers. Our findings extend the knowledge on the consequences of customer waiting, and enable service providers to better understand the financial and operational impact of waiting-related decisions in service settings.

1. Introduction

In the U.S. approximately 37 billion hours are spent on waiting in physical lines annually (Stone, 2012), which adds up to a wait between two and three years in the lifespan of an average American (Cox, 2005). This waiting takes place at a variety of service settings, such as restaurants, banks, amusement parks, retail stores, and healthcare facilities. Waiting for treatment in a healthcare facility might be unavoidable because of the lack of alternative options. In other service settings, however, customers are apparently consciously choosing to spend substantial amounts of time in line before they are served. Even though companies do not directly experience the costs of the discomfort incurred by their customers because of waiting, it is not clear to what extent these costs could have a direct and delayed impact on profitability through customer decisions and actions. In this paper, we empirically investigate several of the implicit consequences of letting customers wait, and we estimate the impact of these consequences in various scenarios using simulation.

The importance of waiting in service practice is to a large extent reflected in the attention academia has devoted to the topic from different perspectives. From an operations perspective, waiting is commonly modeled as a cost function in which the wait results from a mismatch between demand and capacity that could be fixed by tweaking operational parameters (Osuna, 1985). Actual and perceived waiting can then be influenced by capacity, layout, and service and processing policy decisions (Luo et al., 2004; Nie, 2000). A large number of studies focus on the behavioral consequences of waiting by showing that long queues can impact aspects such as service evaluations and customer satisfaction (Davis and Maggard, 1990; Houston et al., 1998; Taylor, 1994), the perceived value of products and services (Debo et al., 2012; Koo and Fishbach, 2010; Kremer and Debo, 2015), and customer loyalty (Bielen and Demoulin, 2007; Dube et al., 1994). At the same time, empirical research and data collection in this domain is challenging. Whereas virtual queueing settings such as call centers are characterized by hi-tech environments in which data is abundantly available (Koole and Mandelbaum, 2002), studies involving physical queues primarily make use of survey data and self-reports (Munichor and Rafaeli, 2007; Rafaeli et al., 2002). These (repeat) purchase

* Corresponding author.
E-mail address: j3.de.vries@vu.nl (J. De Vries).

https://doi.org/10.1016/j.jom.2018.05.001
Received 5 April 2017; Received in revised form 1 May 2018; Accepted 8 May 2018

intentions do not necessarily lead to actual behavior and corresponding capacity usage (Chandon et al., 2005). In the current study we circumvent this limitation by using data on actual customer behavior.

Furthermore, the majority of studies on waiting time and its consequences make several assumptions that might not hold across practical settings. For example, the arrival rate of customers is commonly treated as an exogenous, fixed parameter (e.g. Hwang et al., 2010; Roy et al., 2016). Whereas this might be reasonably accurate if the model describes only a short period of time, this assumption does not hold in the long-term. In reality, a customer who faces an excessive waiting time during a visit to a service provider may renege or never return after leaving unsatisfied after being served. Very few studies incorporate relations between the waiting time and arrival rate at a later point in time (Ittig, 1994; Umesh et al., 1989), which can substantially influence revenue and profit (Ittig, 2002).

Additionally, only few studies acknowledge that the experienced waiting time or the tolerance for waiting might impact customers' service requirement and duration (Wu et al., 2018). Also the service staff has some discretion in determining completion time (Hopp et al., 2007). They might increase their effort and decrease the service duration in response to a slightly higher workload, but they also might become demotivated and unproductive in response to excessive workloads (Tan and Netessine, 2014). This implies that observed customer behavior is simultaneously influencing and influenced by waiting time. Consequently, a service operation should not focus on minimizing the waiting time, but on maximizing revenue through minimizing the costs associated with waiting (Gavirneni and Kulkarni, 2016). To truly understand the operational implications of these dynamics, we combine an empirical model to investigate the isolated consequences of waiting with a simulation model incorporating the combined effects of waiting time on reneging, customer returns, and revenue. This combination between empirical analyses and simulation enables us to experimentally investigate waiting time in the context of specific restaurant policies, and provides results that are better generalizable and meaningful to practice.

More specifically, in this study we aim to address the following research questions:

RQ1. What are the isolated effects of waiting time on customer behavior (in terms of reneging, dining duration, and returning)?

RQ2. What are the dynamic consequences of the combined empirically identified effects of waiting time?

RQ3. How do specific proposed operational strategies, such as increasing capacity, flexible seat allocation, and encouraging reservations impact waiting time and its consequences?

To achieve this, we employ operational field data obtained from 94,404 groups of customers of a restaurant operation. To answer RQ1 we use isolated empirical models to identify the effect of waiting time on customer behavior. To address RQ2, we embed the empirical models in a simulation framework to capture the complex interactions and dynamics of the investigated constructs. This simulation model enables us to demonstrate the (longer-term) impact of waiting time on customer returns and revenue by incorporating the endogenous effect of waiting time on reneging, dining duration, and future arrivals. Subsequently, to address RQ3, we leverage the integrated simulation model to evaluate the effect on revenue of various operational policies that restaurants could deploy. As a consequence, this study should not only help to improve understanding of the waiting process, but could also lead to new insights regarding company policies in order to maximize revenue.

The study is therefore divided in two parts. We first develop hypotheses, explain the method, and test the hypotheses in Sections 2, 3, and 4, respectively. We show the impact of waiting time on return behavior, reneging, and dining duration. We then investigate the impact of several hypothetical operational scenarios on waiting time, return behavior, reneging, and revenue through simulation. This simulation model and the results of the simulation are explained in Section 5. Section 6 draws conclusions and discusses implications for Operations Management theory and practice.

2. Hypothesis development

2.1. Waiting time

Waiting is in many cases one of the first interactions between service providers and customers. Because of this, adequately managing the waiting time is a vital issue (Davis and Heineke, 1998). Waiting time can be considered in a subjective way as the waiting time perceived by the customer, or in an objective way as the actual waiting time. Even though the actual waiting time might differ from the waiting time perceived by the customer, actual waiting time is still the most important predictor of perceived waiting time (Dabholkar, 1990; Thompson et al., 1996). This study therefore focuses on the impact of actual waiting time on three outcomes: customer loyalty, reneging, and dining duration.

2.2. Customer loyalty

For companies operating in competitive markets, obtaining a base of loyal customers is essential for survival (Srivastava et al., 1998). Customer loyalty, which can be defined in terms of repurchase behavior (Estelami, 2000), repurchase intention (De Ruyter and Bloemer, 1999), or long-term commitment to repurchase (Ellinger et al., 1999), can increase profits through reducing the costs associated with acquiring new customers, through generating a base of customers that is less price-sensitive, and through lower operational costs due to the familiarity of customers with the procedures and systems of the company (Hallowell, 1996).

Customer loyalty is especially important in industries with low switching costs for consumers, as consumers can freely decide to move their business to competitors (Shapiro and Varian, 2013). In service contexts such as restaurants, a dissatisfied customer will face virtually no barriers to dine somewhere else next time. One of the most important drivers of customer loyalty is service quality (Devaraj et al., 2001; Stank et al., 1999). The literature on service quality highlights two critical components: relational elements and operational elements. Relational elements refer to activities focused on understanding the needs and expectations of customers. The importance of relational elements of service quality in determining customer loyalty have been demonstrated frequently, mainly in the marketing literature (e.g. Bell et al., 2005; Crosby et al., 1990; Payne and Frow, 2005). Operational elements, referring to all activities service providers perform to achieve consistent high level of productivity, quality, and efficiency (Stank et al., 1999), are essential determinants of service quality as well (Harvey, 1998). Waiting time is such an important operational element of service quality. An increased queue length can attract customers on the short term by signaling quality (Debo et al., 2012; Kremer and Debo, 2015; Veeraraghavan and Debo, 2009), but this effect is only expected to apply in case of quality uncertainty and in case alternative options are available. In deciding whether or not to come back to a restaurant, customers take the actual experienced quality into consideration. A longer wait during a past visit is therefore not expected to make customers more likely to come back soon in the future. Consequently, we expect that waiting time will have a negative impact on customer loyalty, as defined by the time until a customer returns:

H1. A longer waiting time will be associated with longer time until a customer returns

2.3. Reneging

In addition to the longer-term effect of waiting on customer loyalty, waiting time can also have more direct implications. When customers enter a queue, they might observe or be informed about information on the expected length of delay. Subsequently, customers can make a decision between entering the queue or leaving before even joining the queue (balking). Even if customers choose to join the queue it is still possible that they will not wait until being served. During the wait, customers might renege on their decision and leave the queue.

Since balking takes place before even entering the queue, it is difficult for companies to observe the exact share of customers engaging in balking. Reneging can be observed more easily in physical as well as digital queues, and therefore serves as an important performance measure for most revenue-generating service systems (Garnett et al., 2002). A substantial fraction of reneging customers will not return to the service provider. Consequently, it is necessary to explicitly incorporate customer abandonment in models that aim to provide implications for operational decision making (Dai and He, 2011). Mathematical queueing models can be used to estimate the impact of balking and reneging in service, but doing so requires an accurate understanding of human queueing behavior in specific service contexts that can only be obtained through empirical analyses (Batt and Terwiesch, 2015).

Nowadays service providers often provide offline waiting options to supplement or replace physical queues. Examples of offline waiting applications include restaurants that use a buzzer, phone, or text customers to inform them of their waiting status, and call-centers that provide a call-back option to customers (Kostami and Ward, 2009). Offline waiting enables customers to freely engage in other activities, which can make the waiting more pleasant. At the same time, the introduction of offline waiting has increased the relevance of reneging because of multiple reasons. First of all, reneging occurs more frequently because customers might simply forget that they are waiting to be served. Secondly, in physical queues, reneging is directly visible to other customers and the personnel of the service provider. In offline queues, reneging is often only discovered once a customer does not show up when being notified. This implies that customers waiting in an offline queue only possess an overestimate of their queue position, which makes it more likely that they will renege (Jennings and Pender, 2016). For the current study, which focuses on an offline queueing context, we therefore expect that waiting time will positively relate to reneging.

H2. A longer waiting time will increase the likelihood that a customer reneges

2.4. Dining duration

Since waiting time usually results from a temporal mismatch between demand and capacity, the time customers spend in a service process plays an essential role in influencing how long people have to wait before being able to enter the process. After all, customer sojourn time, or dining time in this case, directly impacts capacity. Even though quality cannot be treated independently from service time in most contexts (Anand et al., 2011), the restaurant industry has been trying to reduce the dining durations of customers in order to increase seat turnover (Kimes et al., 2002; Thompson, 2009). However, while the role of dining duration as a tool to impact revenue has been established (Kimes et al., 1999; Kimes and Thompson, 2004), much less is known about what predicts dining duration. The rare examples of studies in this domain include Kimes et al. (2002), who established that Europeans preferred significantly longer dinners than North Americans and Asians, and Kimes and Robson (2004), who demonstrated the impact of table characteristics on dining duration. Because of the clear implications of dining duration on capacity management and resource allocation, advancing the understanding of this topic from an Operations Management perspective is essential.

In this study, we therefore aim to contribute to this literature by studying the role of waiting time in predicting dining duration. From an operational perspective, we expect that two opposite effects could play a role: on the one hand, a long waiting time suggests that a system is running at full capacity. This could lead to slower service once a customer has actually entered the system, resulting in a longer dining duration. On the other hand, in their quest to reduce waiting time, servers and kitchen staff could exert increased efforts and realize a shorter dining duration (e.g. Kc and Terwiesch, 2009; Shunko et al., 2017; Tan and Netessine, 2014; Wang and Zhou, 2017). From the perspective of the customer, two other opposing effects could play a role: on the one hand, customers who have waited for a long time might want to dine longer to make the wait worthwhile. On the other hand, customers only have a limited amount of time available. After a substantial amount of time has been spent waiting, customers have a limited amount of time left for dining. We expect that in the modern society, where time is limited and highly valued (Leclerc and Schmitt, 1999), the impact of waiting time on dining duration is mainly driven by the amount of time a consumer has available for dining. Even if service could be slower or if a longer stay would be desirable after a long wait, we expect that customers will adapt their choices of food and dining speed because of their constrained time (Jabs and Devine, 2006). Customers do not only consider the money spent on dining, but also incorporate the perceived value of their time as part of the perceived price (Becker, 1965; Zeithaml, 1988). This results in the expectation stated in hypothesis 3:

H3. A longer waiting time will shorten the dining duration of a customer

Fig. 1 shows a conceptual model including the predicted empirical relations between waiting time and the outcome variables, as well as the proposed effects of waiting time on customer returns that will be demonstrated using simulations. Furthermore, the conceptual model also shows a positive relationship between dining duration and customer waiting time, as an increase in dining duration impacts resource utilization (tables) and hence increases waiting time of subsequent customers.

3. Methodology

3.1. General approach

The aim of this study is two-fold. First, we use empirical models to test our hypotheses, and to identify the isolated effects of waiting time on several outcome variables. Second, we incorporate the empirically-established findings in a simulation analysis to estimate the system-wide dynamic impact of the empirically-identified effects and analyze

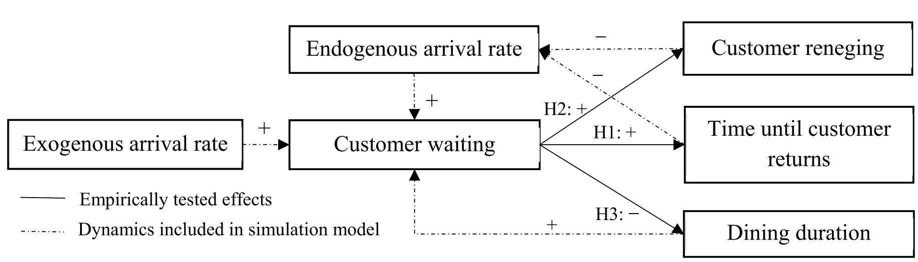

Fig. 1. Conceptual model showing relations between variables.

various scenarios. Using a simulation in addition to the empirical analyses is necessary to simultaneously account for the stochastic nature of multiple determinants of waiting time, and to estimate the impact of the dynamic interactions resulting from the feedback loops displayed in Fig. 1. With some additional model assumptions, analytical modelling approaches such as customized queuing network or system dynamics based models could achieve a similar goal (see Roy et al., 2016). However, analytical models are quite sensitive to the distributional assumptions on the customer inter-arrival times and dining times, and can result in additional approximation-related error in the output measures. These errors can affect the direction of the numerical insights generated from the model. In our case, it is important to accurately describe the real restaurant, and a simulation is better able to handle the empirical data and dynamic interactions with limited model assumptions, and relatively easy to validate.

As a result, the use of a simulation in combination with empirical data in this context enables us to bridge the gap between academic rigor and practical relevance by providing more comprehensive estimates of the impact of specific operational decisions (Shafer and Smunt, 2004).

3.2. Data collection

The data used in this study were collected between February 1, 2016 and January 31, 2017 from a popular dine-in restaurant in Bangalore, India, that uses a sophisticated digital restaurant reservation and table-management platform. The restaurant is located in an area with a high restaurant density. It offers a wide variety of food and drinks, with prices in the upper segment of the local market. The restaurant opens at noon every day and closes around midnight, depending on when the last diners finish their meals. The diners are typically working professionals. We selected the restaurant for this study because of the high share of walk-in customers (above 90%), who could experience waiting. The customers arrive at the restaurant either with a prior reservation or as a walk-in customer. The restaurant has 84 tables with in total 335 seats, and allocates maximum 10% of total capacity to reserving customers. Reserving customers reserve a table using either the website, a mobile application, via phone, or, occasionally, on site. The customer's cell phone number serves as his or her customer ID. The restaurant managers can see all incoming reservations at a glance (see Fig. 2a), and then confirm them using the tablet. A confirmation email and an SMS are automatically sent to the customer.

All customers are checked-in upon arrival using a tablet at the front desk by providing their contact ID. Walk-in customers also provide information about their group size. Based on historical customer data, the system recognizes the customer's visit count to this restaurant. Upon registration, they also receive an SMS with their current queue position (see Fig. 2b) and a website link to check their queue status. They can then either leave the restaurant or wait in or nearby the restaurant for their turn. The automated queue management software updates the diner whenever the queue position changes and sends an SMS when the queue position becomes number one. The tablets are located on all floors where supervisors open up tables. Information is directly sent to the front desk when a table opens up, and the customer is informed via cloud telephony. If the customer group does not show up in 5 min, the table allocation is cancelled. Once the customer group arrives at the front desk, they are guided to their table.

Customers not showing up or canceling after making a reservation are not included in the data. The data obtained from these groups include information on the customer ID, type of transaction (reservation or walk-in), visit number, group size, status (seated or not), time of making a reservation and reserved time (for reserving customers), arrival time, queue position at the time of arrival, seating time, and table number. These data were used to construct our dependent and independent variables. It should be noted that only one person of the group registers in the digital platform of the restaurant. Therefore, individual customers, who are part of a group, serve as the unit of analysis in this study. In our discussion, we use the terms customer and customer groups interchangeably. Table 1 provides an overview of the relevant collected data and corresponding collection modes.

3.3. Operationalization of dependent and independent variables

3.3.1. Return status

For every visit of a customer, as identified by the customer ID, we registered two things: whether he or she returns to the restaurant within the timeframe of the dataset and the number of days until the next visit or the end of the dataset (if no revisit is recorded in the dataset). Tracking the number of remaining days until the end of the data collection period is required to appropriately handle the censored nature of the dataset in a survival analysis (Hosmer and Lemeshow, 1999). For the investigated restaurant, the average return time of returning customers appeared to be around 73 days. Because only one person per group of customers is required to register, it is possible that during subsequent visits another person of the group registers. This would lead us to erroneously conclude that the particular visit was no returning visit. However, there is no reason to expect that the change in registering customer is related to waiting time or any of the other control variables. The empirically identified effects on the return behavior identified in this study should therefore still hold, but the identified return rate might be a systematic underestimation of the true return rate.

3.3.2. Reneging

Every walk-in customer group entering the queue but not showing up when it is their turn is considered as a case of reneging. This occurs in total 14,585 times, corresponding to 15.7% of all walk-in customer groups. For the large majority of the customer groups it is unknown when they left the queue, but 3372 customer groups actively indicated in the system that they were abandoning the queue.

3.3.3. Dining duration

As checkout time of customers was not directly measured, the dining duration was estimated using table occupation. In case a queue was present, we estimated dining time as the difference between seating time and the seating time of the next customer group at that particular table. We subtracted 10 min to account for the time needed to vacate and clean the table. To investigate hypothesis 3, we excluded customers with a dining duration shorter than 30 min (time from being seated until leaving) to ensure that the sample does not include customer groups who decide to leave before ordering. The remaining number of observations of which we have a measure of the dining duration is 35,163. This measure was compared with an estimate of dining duration based on the time at which a subset of 2586 customers provided feedback scores during the payment procedure. A comparison between these two measures revealed only very small differences ($<$ 5%), which suggests that our estimate of dining duration is reasonably accurate.

3.3.4. Waiting time

The waiting time of customers was measured as the difference between arrival time at the restaurant and seating time. It should be noted that in this particular case, customers are not required to stay in a physical queue. Because they are notified when they are first in the queue and once a table becomes available through their phone, customers can engage in other activities while waiting, as long as they stay in the vicinity of the restaurant. Many customers take this opportunity to go e.g. shopping in close-by stores.

For most of the reneging customers (75%), we do not precisely know when they left the queue. As a consequence, we do not know the precise waiting time experienced by these customers before they decided to renege. To address this issue, we employ the initial queue position of these customers as a proxy for waiting time. For the customers with a known waiting time, a strong linear correlation between

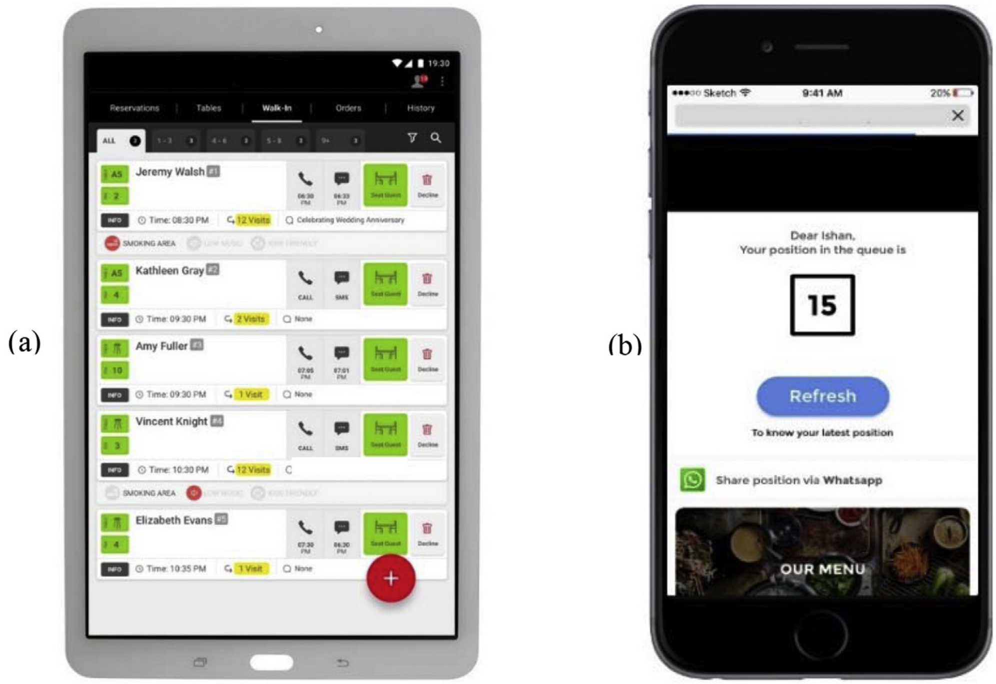

Fig. 2. Image of (a) the reservation app in the restaurant, (b) customer information on queue position. Phone numbers and privacy-sensitive data have been obscured.

Table 1
Data collection.

Data	Collection mode
Customer ID	Customer enters in digital platform of restaurant
Visit number	Tracked by digital platform of restaurant
Type of transaction	Tracked by digital platform of restaurant
Group size	Customer enters in digital platform of restaurant
Arrival time and date	Entered by staff in digital platform of restaurant
Seating time	Entered by staff in digital platform of restaurant

initial queue position and waiting time exists, $r(75{,}725) = 0.70$, $p < .01$. Furthermore, the queue position can also be easily implemented in the subsequent simulations. A comparison between the distribution of the (known) waiting time of reneging customers and seated customers reveals substantial differences between the two groups of customers. The average waiting time for reneging customers is 67.6 min (SD = 40.1), while the average waiting time for seated customers is 20.4 min (SD = 30.5). Also in terms of position when entering the queue, reneging customers have an average entering position of 45.6 (SD = 28.6), whereas seated customers have an average entering position of 15.3 (SD = 20.2).

3.4. Operationalization of control variables

We employed several control variables to capture additional variance between visitors: the visit number, group size, evening, and weekend.

3.4.1. Visit number

The unique customer ID of every visitor is used to track the total number of visits to the restaurant. This does not only include visits in our sample, but also all prior visits made since the restaurant started using the digital reservation platform, January 1, 2015. Even though customers in the sample could have visited the restaurant before this date without our knowledge, the impact of these visits that took place more than fourteen months before the start of data collection is expected to be limited. Controlling for visit number is essential, as customers who have visited the restaurant before can already be considered as loyal customers, whereas newcomers are having a first impression of the restaurant. This will have a substantial impact on the influence of the experience during the specific visit on the time until a customer returns.

3.4.2. Group size

Although we collected data from only one customer in every group of visitors, we control for group size, which is measured once customers register at the front desk in the restaurant.

3.4.3. Evening

Customers coming during the afternoon might differ from the customers during the evening. Furthermore, customers during the evening have a lower maximum dining duration, as the restaurant closes around midnight. To facilitate a fair comparison between customers coming during the afternoon and customers coming during the evening, we included a dummy variable with a value of '0' if a customer was seated in the restaurant before 5.30 p.m., and '1' if a customer was seated at or after 5.30 p.m.

3.4.4. Weekend

Similarly, to control for the potential differences in customers and customer habits between weekends and weekdays, we employ a dummy variable with a value of '0' if a customer visits during the week, and a value of '1' if a customer visits during the weekend.

As robustness check we compared the fit (in terms of Akaike's information criterion) between all estimated empirical models with the untransformed count variables 'visit number' and 'group size', and the models with the log-transformations of these variables. The variable 'group size' provided the best fit without transformation, whereas for 'visit number' the log-transformation explains significantly more variance in the outcome variables. Consequently, we control for 'group size' in its original form, and include the log transformation of the 'visit number' in the empirical models.

Table 2
Descriptive statistics of walk-in customers.

	N	%		*N*	%
Visit Number			*Group Size*		
1	62,771	66.49%	1	614	0.65%
2	15,300	16.21%	2	41,351	43.80%
3	6,688	7.08%	3	19,761	20.93%
4	3,510	3.72%	4	14,830	15.71%
5	2,074	2.20%	5	6,691	7.09%
6	1,281	1.36%	6	4,480	4.75%
7 or 7+	2,780	2.9%	7 or 7+	6,677	7.1%
Weekday	54,625	57.86%	*Afternoon*	33,726	35.73%
Weekend	39,779	42.14%	*Evening*	60,678	64.27%
Reneging			*Visit*		
No	81,528	86.36%	Single visit	70,579	74.76%
Yes	12,876	13.64%	Return	23,825	25.24%

	Afternoon			*Evening*			*Overall*		
	Range	Mean (SD)	Median	Range	Mean (SD)	Median	Range	Mean (SD)	Median
Waiting time seated customers (min)	0–242	6.11 (12.93)	0	0–310	29.31 (34.81)	17	0–310	20.38 (30.04)	5
Position in queue	1–50	4.79 (6.52)	2	1–127	22.92 (23.56)	17	1–310	18.96 (23.12)	6
Dining duration (min)	30–649	161.09 (106.86)	139	30–345	86.04 (62.26)	83	30–649	144.39 (86.87)	105
Days between visits	0–364	87.61 (85.01)	59	0–364	90.58 (88.54)	59	0–364	89.47 (87.26)	59

4. Empirical model testing

4.1. Descriptive statistics

Since only walk-in customers are potentially subject to waiting time, data of the reserving customer groups cannot be used to empirically test the impact of waiting time. The descriptive statistics of the walk-in customers (Table 2) show that most customers visit the restaurant for the first time, the most common group size is two persons, and that the average time between subsequent visits of returning customers is approximately 89 days. The difference between the mean and the median of waiting time suggests a positively skewed distribution, which is (to a smaller extent) also the case for dining duration.

Table 3 displays the Pearson correlations between the variables included in the various empirical models used to test our hypotheses: a parametric survival model to predict the time between visits, a logistic regression analysis to predict reneging, and a negative binomial regression analysis to predict dining duration. The control variables 'Group size', 'Evening', and 'Weekend' show a reasonably strong and positive correlation with waiting time. This can be expected, since it is more difficult to find a seat for larger groups and it might be busier during weekends and evenings, all resulting in longer waiting times. In order to test our hypotheses, it is necessary to disentangle the effects of the control variables from the effects of the focal predictor, waiting time. An inspection of the Variance Inflation Factors (VIFs) of the variables of all estimated models reveals VIFs between 1.14 and 9.84, thus not exceeding the commonly employed threshold of 10 (Hair et al., 2010). Furthermore, similar to the procedure followed by Batt and Terwiesch (2015), we compare the standard errors in our full regression-type models with the standard errors in simplified models without interaction terms. The standard errors were small and relatively stable, suggesting that multicollinearity is not a concern.

Table 3
Pearson correlations of untransformed variables.

		Waiting time (min)	1	2	3	4	5	6	7
1	Dining duration (min)	-.34							
2	Customer returning?	-.02	-.01						
3	Customer reneging?	-.07	N/A	.04					
4	Days between visits (or end data collection)	.01	.02	-.41	.01				
5	Visit number	-.01	.00	.17	-.02	-.14			
6	Group size	.11	.10	-.06	-.01	.05	-.02		
7	Evening (vs afternoon)	.35	-.40	-.01	.19	-.01	-.01	.09	
8	Weekend (vs weekday)	.14	.00	.00	.09	.03	-.02	.01	-.20

Note: because bivariate normality cannot be assumed for most variable pairs, no significance values are provided.

4.2. Impact of waiting time on time until revisiting

To test Hypothesis 1, stating that a longer waiting time is associated with a longer time until a customer returns, we use a fully-parametric survival model. This model enables us to incorporate both the return status (returning or not) and days between subsequent visits of customers, and to handle the censored nature of the data (as customers towards the end of the data collection period had less time to return). We choose a fully-parametric survival model because we are interested in obtaining a hazard function that describes the structural relationships between event times and independent covariates, which can be used in the simulation analyses to model the effects of waiting time on customer return behavior (Bender et al., 2005; Melnyk et al., 1995).

To estimate the impact of waiting time on the time until a customer returns, we focus only on seated walk-in customers, since they experience the full length of the wait before dining. We fit the model using nine months of data (56,369 observations), in order to enable a validation of the simulation model with remaining data of seated walk-in customer groups (20,711 observations).

As initial step, we compare a simple parametric proportional hazard model (without predictors) fitted using an exponential inter return time distribution with a base model fitted using a Weibull distribution with the 'Survival' package (Therneau and Lumley, 2017) in R 3.1.3 (R Core Team, 2017). This revealed that a Weibull distribution fits our data significantly better ($\chi^2_{(1)} = 2100$, $p < .001$). Subsequently, we included 'Waiting time', the control variables, and all possible two-and three-way interactions in the analysis. A systematic backward elimination procedure (Zhang, 2016) yielded a final model that fits

Table 4
Empirical models.

	Model 1			Model 2		Model 3	
	Parametric survival model			*Binary logistic Regression*		*Negative binomial regression*	
Subjects	Seated walk-in customers			All walk-in customers		All walk-in customers	
Dependent variable	Time until returning (days)			Reneging (no = 0, yes = 1)		Dining duration (minutes)	
	Estimate	Std. error	Hazard ratio	Estimate	Std. error	Estimate	Std. error
Waiting time (minutes)	.0007**	.0003	.9994			-.0022**	.0003
Position in queue				.3424**	.0162		
Evening (vs. afternoon)				2.5841**	.1154	-.5823**	.0081
Group size (count)	.0916**	.0106	.9278	-.1390**	.0073	.0451**	.0016
Weekend (vs weekday)				1.7912**	.1234	-.2580**	.0078
Visit number (Log)	−1.2455**	.0338	2.7716	-.2705**	.0217		
Evening × Weekend				−1.1107**	.1311	.3403**	.0119
Visit number (Log) × Group size	-.0401**	.0101	1.0333				
Waiting time × Evening						-.0020**	.0003
Waiting time × Group size							
Waiting time × Visit number (Log)							
Queue position × Evening				-.3010**	.0163		
Queue position × Weekend				-.2520**	.0165		
Queue position × Evening × Weekend				.2333**	.0165		
Waiting time × Visit number (Log) × Group size							
Constant	6.9399**	0.0418		−4.6028**	.1140	5.182	.0077
Observations	56,364			66,427		35,163	
McFadden's pseudo R^2				0.187			
Tjur's coefficient of determination				0.172			
Scale parameter (λ) in Model 1/dispersion parameter (θ) in Model 3	.0034**			1		4.0722**	
Shape parameter (v)	.818**						
Akaike Inf. Crit.	221,331			49,071		381,264	

Note: $^{*}p < .05$, $^{**}p < .01$.
Note 2: Even though the models reported in this table only include significant control variables, it should be noted that the significance of the hypothesized predictors does not change when using the full set of controls in all models.

significantly better than a null model ($\chi^2_{(9)} = 8961.18$, $p < .001$), with significant main effects of 'Waiting time', the log transformation of the number of visits, 'Group size', and a significant two-way interaction between 'Visit number' and group size. Using the 'SurvRegCensCov' package (Hubeaux and Rufibach, 2014) we also estimate the hazard ratio parametrization, which is easier to interpret. The hazard function can be expressed as

$$h(t|x) = \lambda \exp(\beta' x) v\, t^{v-1} \tag{1}$$

where t refers to the time, x to the vector of predictors, β' to the transposed vector of regression coefficients, λ to the scale parameter and, v to the shape parameter. This resulting hazard ratio (displayed in the "Hazard ratio" column of Model 1 in Table 4) of 0.9994 for 'Waiting time' implies that a customer group which experienced the average wait of 20 min during their last visit is approximately 1.1% less likely to return at a given day than a group which did not experience any wait ($0.9994^{20} = 0.989$). The significant shape parameter (v) is smaller than 1, which indicates that the rate of customers returning decreases over time. The significant scale parameter (λ) indicates the extent to which the distribution is stretched out and is a function of the predictors (Bender et al., 2005).

To establish that the identified effects are robust and not a statistical artifact caused by the dependence between repeating visitors, we also estimated a parametric shared frailty model. In this model we account for the fact that several observations are clustered within an individual customer with a specific tolerance for waiting. Using the "Parfm" package (Munda et al., 2012) in R (R Core Team, 2017), we predict the number of days until a customer returns using the same predictors as in the original parametric survival model: waiting time, group size, the log of the visit number, and the interaction between group size and the log of the visit number. Again we use a Weibull distribution for the baseline hazard function. The results of the parametric shared frailty model show that accounting for the dependence in the data does not lead to changes in significance relative to the original survival model: all included variables are significant predictors of the time until returning at a .01 significance level. This result provides us with confidence that the outcomes of the empirical survival analysis do not change substantially when taking the dependence structure in account. As another robustness check, we also repeat our original survival analysis on a subsample that only contains every customer group once to ensure the independence assumption is met. The estimates and significance levels resulting from this analysis are nearly identical to the results of Model 1. Because the contribution of a shared frailty model is in this case limited relative to the increase in complexity, we use the estimates of the regular parametric survival model (displayed in Table 4) as input in the simulation.

Additionally, because evidence exists that people perceive time and waiting time logarithmically instead of linearly (Antonides et al., 2002; Zauberman et al., 2009), we followed the procedure outlined by Lind and Mehlum (2010) and also estimated models with quadratic- and log-transformations of waiting time as predictors. Neither the model including the quadratic transformation ($\chi^2_{(1)} = 1.61$, $p > .99$) nor the model including the log-transformed variable ($\chi^2_{(1)} = 2.01$, $p > .99$) provided a significantly better fit, suggesting that assuming a linear relationship is appropriate. Furthermore, as some studies (e.g. Kremer and Debo, 2015) suggest that a bit of waiting can result in a positive effect on purchase intention, we estimated two additional models: a model with a dummy variable indicating if a customer group encountered any wait or not, and a model including a dummy variable indicating that a customer group waited for less than 5 min. A significant coefficient for either of these dummy variables would indicate that the effect of waiting time on return behavior for customer groups who faced no or only a very short waiting differs from the effect of waiting time for the other customers. In these alternative models, no

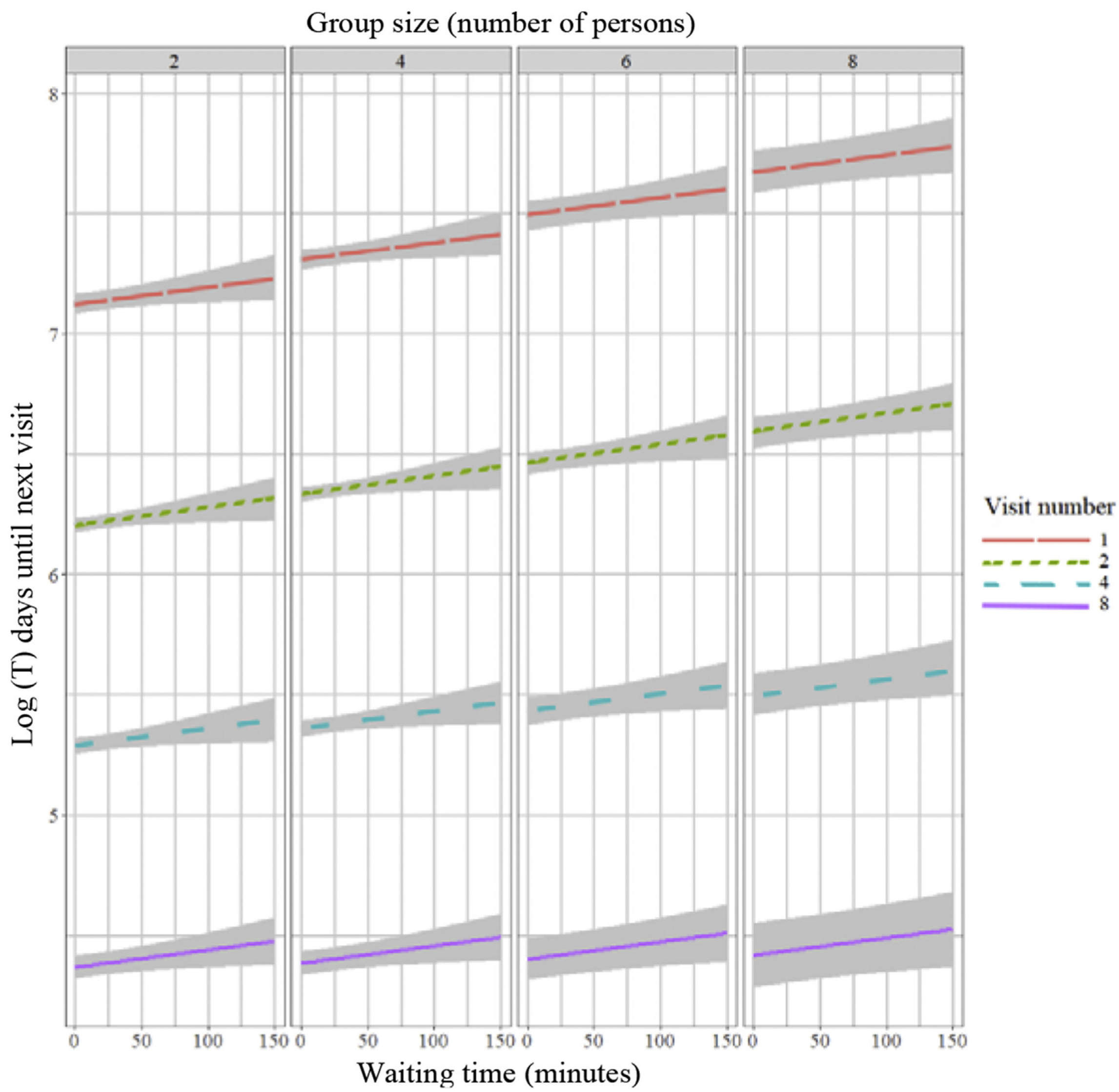

Fig. 3. Graphical representation of the relationship between predictors and the number of days until next visit (Log scale) predicted by Model 1. Ribbons represent 95% confidence intervals.

significant effect of no waiting or little waiting was identified. This suggests that there is no need to model the relationship between waiting and return behavior differently for customer groups who faced no or only little waiting.

Model 1 in Table 4 displays the results of the final parametric survival model used to predict the time until a customer returns. A positive estimate indicates that a higher value of a particular predictor is related to a larger number of days between subsequent visits. The positive estimate of waiting time is therefore in line with Hypothesis 1. However, to properly interpret the findings, it is necessary to take the potential interaction effects into account. Fig. 3 displays the results including the interaction effects. The plot clearly illustrates that the current visit number is an important determinant of future visit behavior, and that the time between visits is substantially higher for larger groups visiting for the first time. For example, a person visiting in a 2-person group visiting for the first time and not experiencing waiting time is expected to return after on average exp(7.15) = 1274 days, a person in a four-person group after exp(7.30) = 1480 days, and a person in an 8-person group after exp(7.65) = 2100 days (see Fig. 3). The visualization of the two-way interaction between visit number and group size shows that the impact of group size on time between visits is smaller when the visit number is higher.

Since reserving customers do not face any waiting time and because we do not know how long reneging customers waited before abandoning the queue, we cannot use the same model to predict their return behavior. For the comprehensiveness of the simulation analysis, we therefore also need to create separate models to estimate the return behavior of these two types of customers. These models, which are shown in Table 8 in Appendix 1, show that 'Visit number' and 'Evening' are significant predictors of return behavior of reserving customers (Model 4). The variables 'Visit number' and 'Group size', as well as the interaction between these two variables significantly predicts the return behavior of reneging customers (Model 5). These results show that 'Visit number' is the most important predictor of return behavior for all customer types. Whether a visit took place in the weekend does not make a difference in terms of return behavior of reneging or reserving customers. The main difference between the various models predicting return behavior is that the variable 'Group size' does not play a significant role for reserving customers, and that the variable 'Evening' does not emerge as significant predictor for reneging customers.

4.3. Impact of queue position on reneging

To test Hypothesis 2, which predicts an increasing reneging probability with an increase in the initial queue position, we need to create a model that predicts whether a customer reneges or not. Several binary choice models are capable of achieving this (Greene, 2012). We tested a logit, probit, and complementary log-log model, all of which delivered similar results in terms of the most important predictors. To facilitate the interpretation and use of the coefficients in the simulation, we present the results of the logistic regression analyses. In the logistic regression model, the variables ‘Group size’, ‘Visit number’, ‘Evening’, and ‘Weekend serve as control variables. Furthermore, we investigate the predictive power of all potential two- and three-way interactions between the control variables and main predictor 'Position in Queue’ to find out whether reneging behavior is dependent on the day of the week and time of the day.

As a first step, we partitioned our full dataset of all walk-in arrivals (90,291 observations) in a ‘training’ dataset containing nine months of data (66,427 observations), and a ‘test’ dataset with the remaining three months of data (23,864 observations). Subsequently, we estimated several logistic regression models based on all our available predictors using the training data, to predict the probability that customers renege. A backward elimination procedure yielded the final model, as displayed in Table 4 (Model 2). To find out whether our model fits the test data well, we performed several tests. A comparison of our model with a null model suggests a good overall model fit, $\chi^2_{(9)} = 11294.68$, $p < .001$. McFadden's pseudo R^2 of 0.187 (McFadden, 1973) and Tjur's coefficient of determination of 0.172 (Tjur, 2009) also suggests a reasonable fit. It should be noted that the relatively small explanatory power of the model (many other factors could affect reneging, e.g. sudden other obligations) does not hamper its contribution, as in the simulations we are interested in the system-wide effects of our predictors rather than distinguishing reneging customers from non-reneging customers with high accuracy. As a next step, we used our model estimates fitted on the training data to predict the reneging probability in the test dataset. The Kolmogorov-Smirnov plot displayed in Fig. 4 shows the predictive power of our model relative to a model without any predictors. The surface under the model curve is 0.799, which shows that our model predicts whether a customer will renege 29.9 percentage points better than the random model. This suggests that our model has a substantial predictive power.

Before interpreting the coefficients, we again ensure that the model results hold when we use a mixed-effects model to account for the dependence structure in the data. Using the “lme4” package (Bates et al., 2014) in R (R Core Team, 2017), we estimate a mixed-effects logistic regression. The resulting coefficients and significance levels are nearly identical to the results obtained without accounting for the dependence structure. We therefore will not use customer-dependent equations to predict the reneging probability in the simulation model.

Model 2 in Table 4 displays the results of the final logistic regression

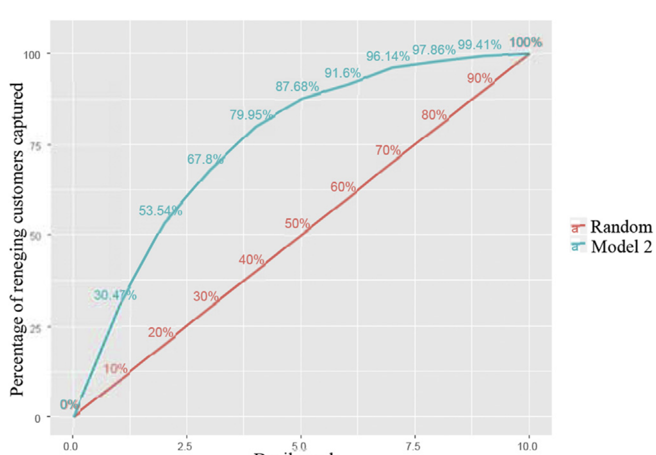

Fig. 4. K-S Plot based on the test data.

model used to predict the reneging probability of customers. The focal predictor ‘Position in queue’ positively relates to reneging probability, whereas larger groups and more regular visitors renege relatively less frequently. The binary control variables ‘Evening’ and ‘Weekend’ also appear to have a significant impact on the reneging probability. Furthermore a two-way interaction between ‘Evening’ and ‘Weekend’ proved significant, as well as a three-way interaction between ‘Position in queue’, ‘Evening’, and ‘Weekend’. Note that coefficients in model 1 are not standardized. Because of the presence of significant interaction effects, we cannot interpret the coefficients of ‘Position in queue’ in isolation. Instead, the impact of waiting on reneging probability is further investigated in (see Fig. 5). This figure graphically displays the combined effect of the queue position and both interacting control variables. The first thing that can be clearly observed in this plot is that a long queue in the afternoon relates to substantially more reneging than in the evening. During a weekday afternoon, customers are almost surely reneging if they enter the queue with twenty-five other groups waiting in front of them. During a weekday evening, the majority of customers waits to be seated even if there are fifty groups ahead of them in the queue. Furthermore, a position further up in the queue is more likely to relate to reneging on weekdays than in the weekend. During an evening in the weekend an increase in queue position of 1 relates to an increased reneging probability of approximately 0.4%, whereas the same increase in queue position relates to an increase in reneging probability of approximately 0.7% on weekdays (see Fig. 5).

In interpreting these findings it is however important to take into account that the average queue length at lunchtime (5 groups) is much shorter than at dinnertime (23 groups), and that in the weekend on average more groups are waiting (22 groups) than on a weekday (12 groups). Consequently, the plots consistently display that the reneging probability increases rapidly if waiting time is substantially higher than the average at a particular moment. To make sure that the use of queue length as proxy of waiting time has similar implications during Afternoon/Evening and Week/Weekend, we checked the correlation between these two constructs in all four subcategories. These correlations were all between 0.65 and 0.75, suggesting that the relationship between queue length and waiting time is stable. The general positive relationship between reneging and queue position supports Hypothesis 2.

4.4. Impact of waiting time on dining duration

We tested Hypothesis 3 by using the ‘MASS’ package (Venables and Ripley, 2002) to perform a negative binomial regression analysis, which was necessary to account for the bounded distribution of the outcome variable ‘Dining duration’ (Hilbe, 2011). A negative binomial model with an estimated dispersion parameter fits significantly better than a Poisson model with fixed dispersion parameter ($\chi^2_{(1)} = 1138707$, $p < .001$). ‘Visit number’ did not significantly impact ‘Dining duration’ and was removed from the model. In building the final model, all possible two- and three-way interaction terms were tested, and two interaction terms emerged as significant predictors of dining duration: the interaction between waiting time and evening, and the interaction between weekend and evening. The resulting model fits significantly better than the null model ($\chi^2_{(6)} = 9439.26$, $p < .001$). Again, we also estimate a mixed-effects model to identify whether the dependence structure in the data impacts our estimates. The resulting model, estimated using the “glmmADMB” package (Skaug et al., 2015), yields highly similar estimates and significance levels. We therefore will not use customer-dependent equations to predict dining duration in the simulation model.

The results displayed in Table 4 (Model 3) show that larger groups of customers tend to dine longer than smaller groups. To interpret the effects of ‘evening’ (vs. afternoon) and ‘weekend’ (vs. weekday) it is necessary to study the significant interaction terms between ‘waiting time’, ‘evening’, and ‘weekend’. Fig. 6 facilitates the interpretation of

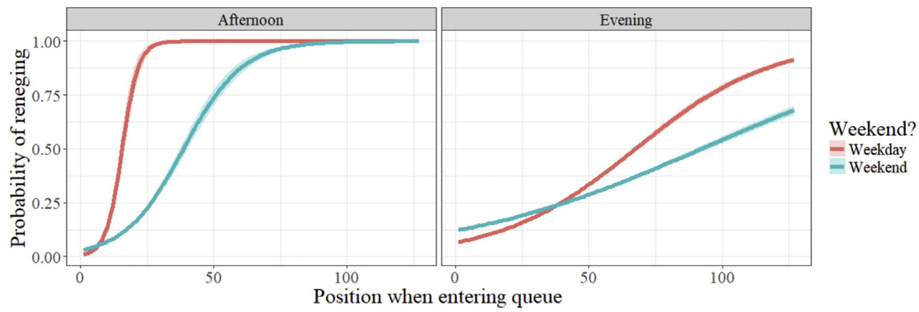

Fig. 5. Graphical representation of the relationship between predictors and the probability of reneging predicted by Model 2. Ribbons represent 95% confidence intervals.

these interaction terms. It shows that meals in the afternoon last longer than meals in the evening. This is partly due to the fact that diners visiting the restaurant during the afternoon have the opportunity to stay longer than diners visiting at night, since the restaurant closes around midnight. We can also observe that in the afternoon there exists a substantial difference in dining times between weekday and weekend visits, whereas this difference is much smaller in the evening. A more important observation is that the effect of waiting time on dining duration is much stronger in the evening than in the afternoon. Specifically, in the evening customers reduce their dining time by approximately 4 min for every 10 min of waiting. Ten minutes of waiting in the afternoon results in a reduction of the dining duration of approximately 2 min.

However, even though this model clearly illustrates the relationship between waiting time and dining duration, it might still be subject to two endogeneity issues: omitted variable bias and selection bias. As we pointed out in the hypothesis development section, dining duration is not only influenced by the preference of the customers. The service speed of the staff can be a predictor of dining duration, and is also potentially influenced by customer waiting time. Since we do not capture service speed, this could be a major omitted variable in our model. To assess whether this is problematic, we investigated with a model that includes the same control variables as Model 3 whether the dining duration of reserving customers (who are never subject to waiting) could be explained by the waiting time of a walk-in group with an identical group size that was seated at a similar point in time (< 15 min difference). Note that the kitchen and waiting staff treat reserving and walk-in customers in the same way. If the speed of service would be an underlying cause of the identified relationship between waiting time and dining duration, we would expect a strong and significant relationship between the waiting time of the 'nearest walk-in group' and the dining duration of a reserving group. However, the negative binomial logistic regression we carried out showed that the waiting time of the 'nearest walk-in group' was no significant predictor ($p = .86$) of the dining duration of reserving groups, suggesting that omitting service speed in our model does not introduce bias.

Furthermore, since our operationalization of dining duration only applies to the groups of customers who finished their dinner while a queue was present, selection bias could potentially limit the generalizability of the findings related to dining duration. For example, it could still be the case that waiting time of a particular customer is not influencing the decision to dine shorter, but that the service provided in the restaurant is simply faster in case a queue is present. Additional robustness tests are therefore required to find out whether the identified relationship between waiting time and dining duration can be generalized to all customers. To investigate this relationship, we used the alternative measure of dining duration based on the time at which a subset of customers provided feedback at the end of their dinner. Since some customers provided feedback while a queue was present and others while no queue was present, this measure facilitates a comparison between the dining durations of both groups. A comparison of the distributions (using a K-S test and plots) revealed no significant difference ($p = .421$) between the distributions of dining durations of both groups, suggesting that the findings should not be subject to selection bias. The consistent evidence that waiting time relates to a shorter dining duration is in line with Hypothesis 3.

5. Discrete-event simulation model

To evaluate the dynamic consequences of the combined empirically identified effects of waiting time and to explore the effect of hypothetical changes in the restaurant policy, we use a discrete-event simulation model. In this model we include the empirically identified consequences of waiting time on reneging, returning, and dining duration simultaneously. As such, the simulation includes some feedback loops: a longer waiting time relates to a shorter dining duration and a higher reneging probability, which shortens the expected waiting time of subsequent customers in line. Similarly, a longer waiting time increases the time until a customer returns, which shortens the expected waiting time of future customers through a reduction of endogenously generated returning customers. Combining both exogenous and endogenous arrivals to estimate the impact on revenue cannot be realized by

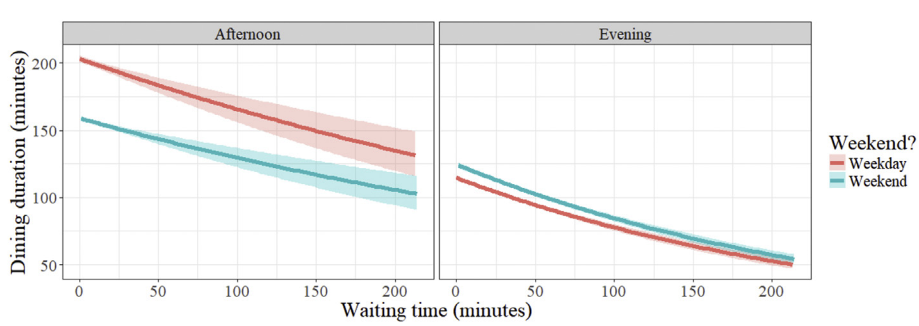

Fig. 6. Graphical representation of the relationship between predictors and the dining duration predicted by Model 3. Transparent ribbons represent 95% confidence intervals.

statistical models such as ARIMA because of the small share of customers that return, and because of the variation in time-lag between returns. A discrete-event simulation enables us to evaluate the potential impact of waiting time on revenue across different scenarios. We used AutoMod™ version 14.0 to develop two model variants. The first variant is a *basic model*, based on all real arrivals (first timers and returning customers). This *basic model* was used to validate the waiting time distribution generated by the simulation model, which demonstrates to what extent we are able to replicate the operational policies employed in the real restaurant. The second variant is the *full model*, which is only based on first time real arrivals and where returning customers were generated by the simulation. This full model was validated using the distribution of returning customers and then used for scenario analysis. Detailed information about the results of the validation and on how specific the restaurant characteristics (number of tables, table assignment, hours of operation, etc.) were employed in the simulation can be found in Appendix 2. The simulation model includes several processes that each customer group encounters at a restaurant: a customer arrival and return process, a customer waiting or reneging process, and a customer dining process. These processes are briefly described in the following paragraphs.

5.1. Customer arrival and return process

The customer groups arrive at the restaurant facility in this process. The basic simulation model uses all empirically observed customer arrivals in the period February 2016–October 2016. See the detailed flow chart in Fig. 7 in Appendix 2. The full simulation model only uses the empirical first-time customer arrivals, and returning customers are endogenously generated by the simulation model. The return time is determined based on empirical Model 1 (in the case of a walk-in customer), Model 4 (for reserving customers), or Model 5 (for reneging customers). Following Bender et al. (2005), these survival models are converted into an expected return time using the following equation (see also the detailed flow chart in Fig. 8).

$$T = \left(-\frac{\log(U)}{\lambda \times \exp(\beta' x)}\right)^{\frac{1}{v}} \tag{2}$$

where U is a random variable drawn from a uniform distribution between 0 and 1, and where the independent variables x, their coefficients β, the shape parameter λ, and the scale parameter v are obtained from the empirical models displayed in Table 4 and 8.

5.1.1. Customer waiting or reneging process

The customer waiting or reneging process starts as soon as the customer group arrives and a table category is allocated to the group. The waiting time experienced by the customer begins at this moment. Furthermore, the queue position upon entering the queue is plugged into empirical Model 2 to determine the reneging probability of an arriving customer group. This reneging probability is compared to a random number drawn from a uniform distribution between 0 and 1. If the reneging probability is higher than the random number, the customer reneges. If the probability is lower, the customer waits until being seated.

5.1.2. Customer dining process

Once a table becomes available from the assigned table category, the customer is assigned to the table and the waiting time ends. The customer then dines at the restaurant. If the customer group is based on an empirically observed arriving group and the dining duration is known, we use the known dining duration in the simulation. If the dining times are not known or if the customer group is an endogenously generated arrival, we use empirical Model 3 (Table 4) to estimate their dining time. The customer waiting time obtained from the simulation model is used as a parameter to estimate the dining time. After the dining duration is over, the customer departs and the table is vacated for use of the next customer group.

The results of the *basic model* validation (Appendix 2) demonstrates that our model accurately simulates the operational policies used in the restaurant, and the out-of-sample validation of the *full model* (Appendix 2) shows that our model is able to able to predict aggregate customer behavior well. Furthermore, the *full model* validation shows that incorporating the dynamic (feedback loop) effects result in estimates that are substantially more accurate than the estimates generated by a model not incorporating these dynamic effects. In the next step, we use the simulation model to assess the impact of several operational scenarios on customer waiting time, which affects the chances of reneging, dining duration, time until revisiting, and hence, restaurant revenues.

5.2. Dynamic effects of waiting time

As mentioned in the description of the simulation model, a longer waiting time leads to several consequences that might reduce waiting time in the short-as well as longer-term. Similarly, a specific exogenous reduction of waiting time is expected to result in a smaller effective waiting time reduction. To explore if an equilibrium situation exists in which the endogenous effect exactly counters the endogenous change in waiting time, we have investigated the effect on average waiting time of exogenously reducing the queue length. To identify the potentially opposite dynamics that take place (exogenous arrivals vs. endogenous effects) we run our simulation model with the queue based on actual arrivals and corresponding waiting time, while using specific discount factors for the waiting time parameters employed to estimate reneging probability, dining duration, and return behavior in the simulation. This approach allows us to compare the waiting time we would expect based on the exogenous reduction of arrivals with the waiting time we estimate using the endogenous effects. The results of this approach indeed show that relatively small exogenous reductions of the waiting time are partially offset by endogenous effects (people eat longer, renege less often, come back sooner). A balanced situation occurs between a 15% and 20% exogenous reduction, where the effective reduction is approximately equal to the exogenous reduction. The results also show that even larger exogenous reductions are in fact reinforced by the endogenous effects. However, it should be noted that especially the results for these larger reductions should be interpreted with care. In this exercise, we pretend that we can extrapolate the empirically identified effects. This assumption is reasonable for small reductions, but results are substantially more unreliable for larger reductions.

Based on these results, we may conclude that while evaluating the costs resulting from customer waiting time, it is necessary to employ a system-wide perspective instead of a component-specific perspective. For example, benefits resulting from a reduction of waiting time, should be compared to associated costs of capacity, customer returns, and goodwill.

5.3. Scenario analysis

We use the full simulation model to explore the relative impact of different scenarios on the impact of waiting on the number of reneging customers, returning customers, and revenue. For each scenario ('base case' scenario, 'reduction of waiting time' scenarios, and 'operational strategies'), we run 15 replications and obtain 95% confidence intervals for our estimates. Since the exact spending per person was not captured, we use the known average spending per customer, depending on group size: \$15.65 per person for one or two-person groups, \$18.78 per person for three or four-person groups, \$21.91 per person for five or six-person groups, and \$15.65 per person for groups of seven or more persons.

5.3.1. Base case

The actual number of customer group visits between February 2016 and January 2017 is 94,404. The input data for the base case scenario,

Table 5
Simulated impact of the effect of waiting on revenue. The numbers between square brackets represent the minimum and maximum value obtained from 15 replication runs.

Feb 2016–Jan 2017	Base case	Half the impact of waiting time	No impact of waiting time
Total # of visiting groups	[93,156–93,719]	[93,083–93,819]	[93,118–94,092]
Total # of returning groups	[22,563–23,119]	[22,487–23,226]	[22,527–23,506]
Total # of reneging groups	[12,364–12,629]	[8357–8659]	0*
Total revenue generated	[$5,033,312 - $5,062,320]	[$5,271,271 - $5,307,804]	[$5,749,495 - $5,811,818]
CI avg. waiting time (min)	8.523 ± 0.098	4.688 ± 0.048	0
Revenue relative to base	–	+4.7%	+14.5%

*In case of no waiting time, we assume reneging does not take place.

Table 6
Simulated impact of operational policies on revenue. The numbers between square brackets represent the minimum and maximum value obtained from 15 replication runs.

Feb 2016–Jan 2017	Extra capacity	Chair pooling
Total # of visiting groups	[93,153–93,986]	[93,077–93,984]
Total # of returning groups	[22,555–23,388]	[22,482–23,387]
Total # of reneging groups	[6118–6282]	[6110–6403]
Total revenue generated	[$5,413,145 - $5,460,937]	[$5,409,652 - $5,443,556]
CI avg. waiting time (min)	6.256 ± .045	5.775 ± .027
Revenue relative to base	+7.7%	+7.5%

in which we considered the empirically identified effect of waiting time on return behavior, reneging behavior, and dining duration, included 70,579 unique customer groups. The simulation run for the base case scenario resulted in an average of 93,451 total visits, including 23,036 return visits. The total revenue generated by the base case scenario is on average $5,047,816, which is only 0.65% lower than the revenue generated in the actual dataset. These results show that our model simulates the actual situation accurately.

5.3.2. Reduction of waiting time

We demonstrate the impact of reducing waiting time on returning customer visits in two scenarios. In particular, in the scenario 'Half the impact of waiting time' we investigated what the expected restaurant revenue would be if the waiting time could be reduced by 50%. The results, displayed in Table 5, show that this would result in an average wait of 4.69 min and a total revenue of $5,289,538, approximately 4.8% higher than the base scenario. Note that the increased revenue is the result of a combination of a reduction in reneging customers and an increase in dining and returning customers. In Table 5, we also show the potential impact of eliminating waiting from the restaurant altogether. For this hypothetical reduction of the waiting time to zero, we also logically assume that no reneging takes place. The resulting average revenue is $5,780,657, or about 14.5% higher than the base scenario. This increase shows that if the restaurant somehow would be able to immediately accommodate all arriving customers, revenue could be boosted substantially. It is important to realize that this result even underestimates the potential of the true effect occurring in practice, as we do not capture customer balking in our model.

5.3.3. Operational strategies

In addition to providing estimates of the potential benefits, we also tested the impact of three specific strategies that the restaurant could employ. The first of those strategies focused on utilizing the backup capacity that the restaurant could employ in case of excessively busy times. More specifically, if the restaurant stretches its existing capacity to the limit, it is able to open two additional tables of each category (14 tables with 76 seats in total), with the same staff. This corresponds to a capacity increase of approximately 20%. The results (Table 6) show that the average waiting time would be reduced from 8.52 min to 6.26 min, and that the total revenue increases to $5,437,041, a 7.7% increase relative to the base scenario. It is important to realize that his increase in revenue takes place with the original arrival pattern, and without attracting any new customers to fill up the additional capacity. Hence, we can conclude that for service providers that frequently operate at full capacity, effectively managing customer waiting time through buffer capacity utilization can generate additional revenue without changing customer demand.

In the second operational scenario, we did not employ the backup capacity but the seats and tables of the original capacity can be deployed in a modular fashion. As such, we treated every seat as a discrete resource, and therefore the restriction that every table can only host one customer group at the same time does not apply anymore. This meant that multiple smaller capacity units can be combined to form a larger table, or split up to facilitate the seating of smaller groups. This additional flexibility in customer assignment allowed us to reduce average waiting time to 5.78 min and to increase revenue to $5,426,603, about 7.5% higher than the base scenario (Table 6). This scenario demonstrates the maximum revenue limit that can be obtained by using modular capacity unit, subject to spatial constraints. Hence, for service providers that frequently operate at full capacity, effectively managing customer waiting time through flexible capacity allocation using modular capacity units provides another way to generate additional revenue without changing customer demand.

In the third operational scenario, we investigate the potential benefits of stimulating customers to reserve a table. We acknowledge that not all walk-in customers are able to make a reservation, but if the restaurant could incentivize a share of walk-in customers to do so (through discounts or the perspective of avoiding waiting time) capacity can be better matched with demand. In the base case (displayed in Table 5) approximately 10% of customers reserves in advance. We compare this cases with cases in which no customers reserve, 30% of customers reserve, and 50% of customers reserve. The results show the revenue gradually increases if a larger share of customers reserves. This makes sense, because a higher share of reserving customers implies that fewer customers will renege. Interestingly, Table 7 also shows that the total number of visiting customer groups decreases if a larger share of customers reserves. This can also be explained by the fact that fewer customers renege, and that reneging customers are more likely to come back sooner. If these customers dine instead of leaving before being seated, they will generate not only more direct revenue but also increase the average time until returning. As part of this third operational scenario, we also investigate the potential gains that can be realized if no customers renege. If the restaurant is somehow able to

Table 7
Simulated impact of the walk-in/reservation mix or absence of reneging on revenue. The numbers between square brackets represent the minimum and maximum value obtained from 15 replication runs.

Feb 2016–Jan 2017	0% reservations	30% reservations	50% reservations	No reneging
Total # of visiting groups	[93,016–93,785]	[91,862–92,520]	[91,028–91,852]	[89,138–89,989]
Total # of returning groups	[22,412–23,184]	[21,267–21,923]	[20,429–21,264]	[18,543–19,396]
Total # of reneging groups	[12,715–12,971]	[8811–9092]	[6203–6461]	[0 - 0]
Total revenue generated	[$4,993,732 - $5,050,511]	[$5,159,910 - $5,209,892]	[$5,293,241 - $5,330,514]	[$5,579,044 - $5,646,740]
CI avg. waiting time (min)	8.316 ± 0.112	9.072 ± 0.083	9.371 ± 0.107	10.712 ± 0.104
Revenue relative to base	−0.5%	+2.7%	+5.2%	+11.2%

avoid that customers renege (for example through reducing the perceived waiting time by offering entertainment and distraction), the restaurant revenue can be increased by 11.2%. Hence, for service providers that frequently operate at full capacity, increasing the proportion of reserving customers can reduce reneging behavior by preventing a mismatch between customer demand and capacity.

6. Conclusions and discussion

Research on longer-term consequences of waiting on performance in service environments is scarce. This does not do justice to the relevance of the topic for customers and businesses, which emphasizes the need for more research in this direction. The current study makes a unique contribution by combining empirical research on consequences of waiting with a simulation model that shows the effects on customer visits and revenue. Using only empirical models, we addressed RQ1 by drawing relationships from historical data and showing how waiting time affects return behavior, reneging behavior, and dining duration. Likewise, by only using the simulation model, we could have observed the effect of fixed parameter setting (with a distribution of dining times) and a fixed arrival stream of customers on the customer waiting time.

However, our integrated approach enabled us to combine exogenous and endogenous arrivals, to handle effects that are non-linear and non-stationary, and to incorporate (higher-order) time-varying interactions. More specifically, to identify the dynamic consequences of the combined empirically identified effects of waiting time, as included in RQ2, we analyzed the effect of customer- and time-varying parameter setting on revenue outcomes of the restaurant. This integrated simulation model considers the empirical relationships such as the waiting time effects on dining duration and reneging probability (through established relationships), along with other dynamic interactions. The empirical relationships guided the parameter settings in the simulation. While the isolated models (empirical and simulation) are instrumental in capturing partial system effects, the integrated model captures the system-wide effects in a more comprehensive manner. The resulting comprehensive model also allowed us to explore the impact of specific operational strategies on restaurant revenues, as posed in RQ3. The results provide several important practical as well as theoretical implications.

6.1. Implications for practice

In the empirical study we established that customers who are subject to longer waiting times are more likely to renege. If they do not renege, customers who experienced a longer waiting time dine shorter, and the time until they return to the restaurant increases. This finding suggests that even though a restaurant manager could be satisfied by seeing a queue of customers waiting for a seat at a particular night, the long-term implications for the restaurant might be less positive. After all, the customer has to decide whether the (expected) service provided is worth the wait. On the short term, the customers in the queue will result in a high occupation rate in the restaurant, and a queue might even provide a positive signal towards new potential customers (Debo et al., 2012; Kremer and Debo, 2015).

A separate question is if the identified impact of the isolated empirical effects is relevant and adding value to the restaurant. To estimate the overall impact on restaurant revenues, we demonstrated in this study that it is necessary to use an integrated model that accounts for the dynamic relationships between the variables we consider. For example, the example of a 1.1% lower chance for a customer experiencing the average evening wait of 20 min does not simply materialize into 1.1% fewer return visits in total. Instead, when fewer groups return, waiting time of other customer groups will be affected at a later point in time, which then affects their reneging behavior and dining duration. The specific combined impact of these variables, and the relative importance of each variable, will be highly dependent on several specific restaurant characteristics.

For example, as the simulation demonstrates, in the focal restaurant a complete elimination of waiting results in a nearly 15% increase in revenue. The generalizable relevance of the effects partly depends on t extent to which the restaurant depends on repeating customers, table categories and the number of tables per category, the availability of alternative dining options, the reservation policy, etc. A restaurant at a touristic location might not expect any returns anyway and might in fact benefit from the signaling value of a queue. However, at touristic locations a queue can be risky as well, as customers facing a long wait might balk or renege. For a restaurant in a local neighborhood, returning customers may be vital for its survival, but reneging might be a smaller problem because alternative options could be unavailable. For both types of restaurants, effectively managing queues and waiting time is essential, but the impact of specific policies can be totally different between cases. Furthermore, it should be noted that in the investigated restaurant waiting is relatively pleasant, because customers do not physically have to stand in line. This means they can engage in other activities while waiting. As a consequence, the identified negative effect of waiting time on reneging and return behavior could very well be larger in contexts where customers are expected to stand in a physical line. Our results demonstrate that ignoring the effect of waiting time on customer reneging, returns, and the subsequent impact on arrival rates can lead to highly unrealistic results in estimating operational performance.

Various approaches could be taken to reduce waiting time in order to decrease reneging behavior and to create more returning customers. In case the marginal costs of additional capacity weigh up against the marginal revenue per customer, capacity can be increased to reduce waiting time during peak periods. Alternatively, differential pricing can be employed to ensure that the customers with the highest reservation price are facing shorter waiting times.

Even though the magnitude of the effect of waiting time is expected to vary across specific contexts, the effect size identified in the current study is likely to be an understatement because we do not capture balking and potential spillover effects resulting from customers informing other potential customers about the long waiting times. Regarding the generalizability of the findings, we expect that the results also apply in other service contexts without admission costs where customers can engage in other activities while waiting, are informed about their queue position, and have sufficient freedom to switch (in terms of proximity and availability) and choose between service providers. The effects on dining duration might be restaurant-specific. However, even though the data for this study was collected in the specific cultural context and among a higher segment of customers, we have no reasons to believe that the effects of waiting time on reneging and return behavior would not apply in different service settings subject to these similar conditions. Examples could include hairdressers, amusement park rides, and take-away/delivery restaurants. Follow-up research will be necessary to confirm whether the magnitude of the identified effects is comparable across these other service-settings, cultures, and customer segments. Less applicable contexts are call center and healthcare queues. Even though these settings also involve waiting, reneging, and return behavior, the limited opportunities for customers to flexibly switch to alternative service providers distinguish these settings from the research context of the current study.

6.2. Implications for theory

Short-term effects of waiting have been studied before in the operations management and marketing literature. Still, several important research gaps remain. First, despite the fact that waiting time has been frequently linked to customer satisfaction (Davis and Maggard, 1990; Houston et al., 1998; Taylor, 1994) and repurchase intention (Bielen and Demoulin, 2007; Dube et al., 1994), the current study is unique in providing clear Operations Management implications by demonstrating to what extent waiting time materializes in actual reneging, altered service requirements, and return behavior through empirical analyses and simulation. The combination of these two methods is of pivotal value in the current study. The empirical models alone rigorously describe the isolated consequences of waiting time, but it is impossible to come up with estimates of the system-wide impact of waiting in practice and to explore how specific operational strategies interact with the waiting-related dynamics. A simulation model alone is suitable to model the basic dynamics of the service process, but does not fully generalize to practice without adequately incorporating the empirically identified waiting-dependent customer behavior. As such, we contribute to the existing literature on waiting and its consequences in multiple ways.

For example, the empirical results demonstrate that part of the heterogeneity in customers' service requirements can be explained by the waiting time they experienced (Gavirneni and Kulkarni, 2016; Wu et al., 2018). This means that waiting time influences the operational policy optimal to serve the customers who did not renege. In the case of the investigated restaurant we observed a negative correlation between waiting time and dining duration. In line with Wu et al. (2018), we demonstrate that such a correlation can substantially impact throughput and the resulting revenue.

Beyond the immediate change in service requirement caused by waiting, the simulation enabled us to demonstrate the longer-term impact of waiting through future arrivals. Typical queueing models in operations management do not consider the potential effect that an increase in capacity might result in an increase in customer arrivals (Ittig, 2002). Our results show that including current waiting time as a predictor of future demand (Ittig, 2002, 1994) leads to more accurate estimates of the arrival rate, which has clear implications for capacity management in service settings with a substantial dependency on repeating customers.

Third, the unique dynamic combination of empirical models in a comprehensive simulation of the restaurant enabled us to demonstrate that the restaurant can realize substantial gains in revenue through increasing the capacity, allowing more flexible allocation rules, or stimulating customers to reserve. Having customers reserve essentially enables the restaurant to differentiate between flexible customers, practically indifferent between multiple service providers, and dedicated customers, with a strong preference for the specific restaurant (He and Chen, 2018). Maglaras et al. (2017) and Afeche (2013) suggested that this differentiation can be exploited if different pricing structures or even strategic delays are used to differentiate customer classes, even more so if the less-delay sensitive customers are more price elastic. Our results suggest that the benefits of such customer segmentation might be even larger when the longer-term effects of waiting or reneging on customer returns are considered as well.

At the same time, the potential gains in revenue would be even bigger in service contexts characterized by a positive correlation between waiting time and service duration (Wu et al., 2018). As such, the results of the simulation analyses reinforce Gavirneni and Kulkarni's statement that it is essential to focus on minimizing waiting-related costs rather than on minimizing waiting time (2016). One promising avenue through which the waiting-related costs can be decreased capacity is capacity sharing. This is more common in manufacturing settings than in hospitality services, but could be beneficial for restaurants under specific conditions. For the focal restaurant, which is characterized by clear peaks of visitors in weekends and in evenings, this would entail teaming up with a neighboring restaurant with similar work content but facing different demand peaks (Yu et al., 2015).

6.3. Strengths, limitations and avenues for future research

The combination between the use of empirical analyses on unique transaction data and simulations to investigate impact on operations and revenue is a vital strength of this study. At the same time, several limitations still exist. For example, the empirical models are subject to the assumption that all observations are independent. Since we propose that waiting time relates to return behavior, and return behavior relates to waiting time, this assumption cannot be strictly maintained in our model. However, since waiting time will only influence a fraction of return behavior (in addition to visit number, group size, etc.), and returning customers make up only a fraction of all arriving customers, the impact of waiting time on future waiting time is only a fraction of a fraction in our empirical models. Consequently, treating the customer arrivals as independent observations is not threatening the validity of the findings. Still, in order to estimate similar models in contexts characterized by a substantially higher dependence on returning customers (and a low exogenous arrival rate), it is certainly necessary to actively take this threat into account.

Another limitation is that we were only able to use the customer ID of one customer in every customer group. Even though we control for group size to at least partly mitigate this limitation, a visitor might also change phone number in the meantime. This would mean that our estimate of the customer return rate is probably an underestimation of the true rate and in reality the impact of the identified effects might be even larger.

Also, we cannot deduce from the data at which exact point in time reneging customer groups exactly abandoned the queue, and assume in the simulation model that the reneging decision is revealed after they have experienced the wait. In reality they will abandon the queue earlier, and influence the remaining customers by doing so.

Furthermore, the models we estimated and the corresponding simulation only consider customers who actually entered the queue or were directly assigned to a seat. This means that we are not able to draw any conclusions about the behavior of customers who balked and decided not to enter the queue. For example, we currently do not know to what extent queue length influences balking behavior, and to what extent balking influences return behavior.

Also, we did not control for possible restaurant promotions or changes in the menu offerings during the data collection period, which might mitigate the negative effect of waiting. Follow-up studies could include the potential impact of vouchers to encourage customer returns, or the potential effect of peak pricing policies. An interesting next step would also be to investigate whether the negative effect of waiting time on return probability is mediated by lower customer experience evaluations.

Acknowledgements

We acknowledge the support extended by the restaurant and data management team in performing this research. Furthermore, we are very thankful for the constructive comments of the reviewers and editors. We believe that the paper has substantially improved based on their valuable feedback.

Appendix 1. Return behavior of reserving and reneging customers

Table 8
Survival models predicting return behavior of reserving and reneging customers

	Model 4		Model 5	
	Parametric survival model		*Parametric survival model*	
Subjects	Reserving customers		Reneging walk-in customers	
Dependent variable	Time until returning (days)		Time until returning (days)	
	Estimate	Std. error	Estimate	Std. error
Evening (vs. afternoon)	0.196*	0.0958		
Group size			0.0961**	0.0198
Visit number (Log)	−1.604**	0.0642	−1.3824**	0.0735
Visit number (Log) × Group size			−0.0447*	0.0226
Constant	7.518**	0.1086	7.07868**	0.0726
Observations	3020		9776	
Scale/dispersion parameter (*v*)	1.25		1.27	
Shape parameter (λ)	.002		.004	
Akaike Inf. Crit.	11,321		38,767	

Note: $^{*}p < .05$, $^{**}p < .01$, the values between brackets represent standard errors of the estimates.

Appendix 2. Simulation protocols and validation

Protocols for the basic and full model simulations

The models simulate the customer arrival and dine-in process for the restaurant. In each model, each category of table is modeled as a multi-server resource. Deciding the restaurant operating hours, number of tables to be opened in each category, the rule for assigning a customer group to a table category, and the process to estimate the dining time for a customer group are described in the simulation protocols. Note that we do not model waiter resources explicitly because we account for the delays in service and other resource congestion effects in the dining time estimates.

Time of operation: The simulation model for the restaurant assumes that the restaurant opens at 12 noon every day and accepts customer arrivals until midnight. The time difference between the last customer arrival of the day and the first customer arrival from the next day ensures that the operating hours of the restaurant is adhered. This time difference is reflected in the simulation by using the appropriate inter-arrival time between the last customer arrival of the day and the first customer arrival of the next day. During this long inter-arrival time, all customer groups from the previous day are cleared from the restaurant.

Number of tables in each category: We analyzed the number of tables that are made available in each table size category during one year. The number of tables that are opened for customer seating does not only vary according to the day of the week, but also varies across afternoon and evening hours on each day. To appropriately handle this, we collected descriptive statistics for the number of tables that are used on each day and also per time interval of the day (afternoon versus evening) for each table category. In the simulation, we set the capacity of the tables based on the 75th percentile value of the number of tables used on a particular day of the week and part of the day. For example, while the number of available tables with capacity four is 35, only 31 tables are opened during Sunday evening hours for 75% of the times. In contrast, only 25 tables are opened for seating during Monday evening hours.

Table assignment: We analyzed the historical assignment of the category of table to a particular customer group and obtained the distribution of table categories assigned to customer groups. For example, we found that tables of capacity two, three, four, five, and six were assigned to customer groups of two persons during 66.37%, 12.35%, 19.13%, 0.20% and 0.12% of the occasions, respectively. We adopted these distributions to randomly assign the category of table to customer groups. Note that the larger capacity table groups may be assigned to a customer group if lower capacity tables are unavailable; however, it may also be assigned to the customer group if the customer group specifically requested a larger capacity table. Since we are unable to distinguish the cause of assignment, we use a random assignment of tables based on the estimated probabilities.

Basic model description

In this model, the customer arrival times are based on real restaurant data. The real data indicates if the customer group is seated, and may include multiple revisits of the same customer group. The fields present in the input data file are shown in Table 1.

Each table belonging to a table category is modeled in the simulation as a discrete server. Hence, each table category is modeled as a multi-server system. Customer groups are first matched to a table category based on the assignment rule. The customer group then joins the queue that belongs to its assigned table category. The tables are assigned to customer groups using a FCFS scheduling rule. Note that the size of the customer groups waiting in the same queue could be different. If the customer group finds an available table from its assigned category, the group is matched immediately and no customer waiting for a table occurs. Else, the customer group waits in the queue for an available table. Customers may also depart without being seated. When a customer group departs after having dined, we allow 5 min to prepare the table for the next group. The waiting time of a customer for an available table is recorded to obtain the distribution of waiting time. The simulation process for the basic model is shown in Fig. 7.

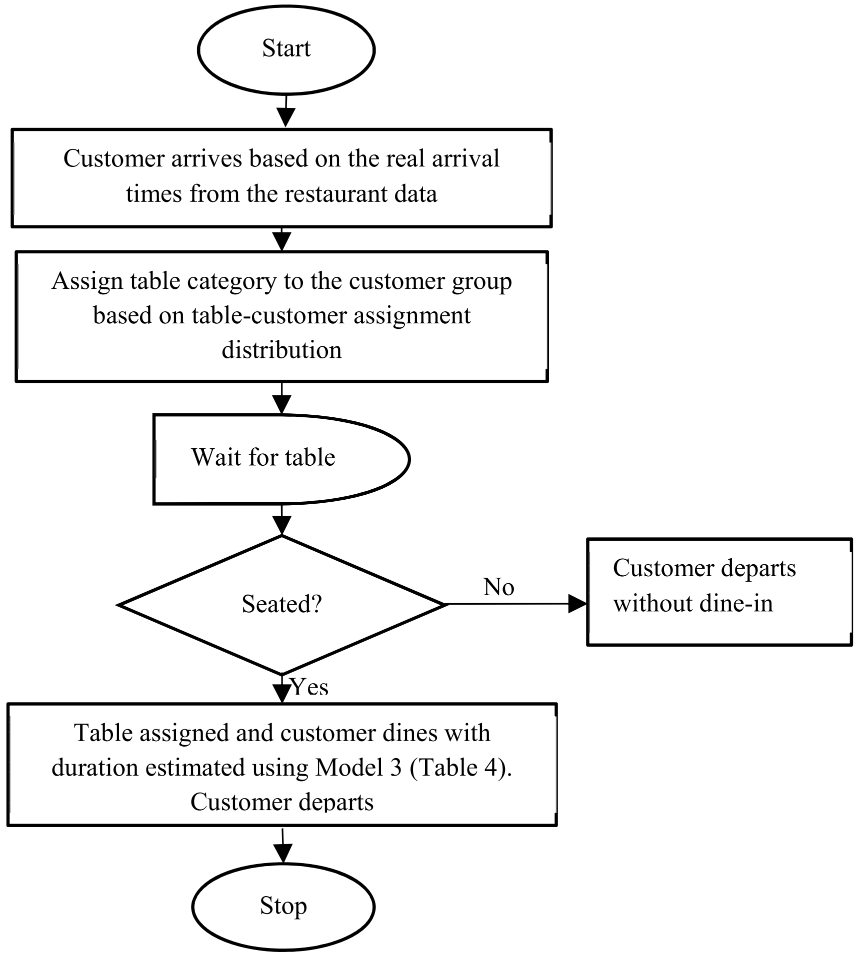

Fig. 7. Basic simulation model, with steps followed for each customer group.
Note that the restaurant may adopt ad-hoc policies leading to extra flexibility that we do not capture in the simulation model, such as merging tables of lower capacity to accommodate seating capacity for larger group sizes. Further, we do not account for any operational delay due to staff unavailability, or consider other sources of variability in the system such as equipment breakdown. Due to the flexibility in seating arrangements in real-life and the actual number of tables in a particular category opened for dine-in, we expect our model to introduce some errors in estimates of the customer waiting times observed at the restaurant for the customer groups.

Full model description

The full simulation model is an enhanced version of the basic simulation model. In contrast to the basic model, where we had a single customer arrival source (real restaurant data) and we considered only the customer arrivals for all customer visits present in the real restaurant data, the customer arrivals used in the full model are combined from two sources: real restaurant data for first time arrivals (exogenous) and simulation-based for returning customer arrivals (endogenous).

In the full simulation model, the empirical models are embedded for estimating the customer reneging probabilities and return times. The logistic regression model developed in the empirical part of the manuscript (Model 2 in Table 4) is used to predict the reneging probability of a customer group, and the survival models (Model 1 in Table 4 and Model 4 in Appendix 1) are used to simulate the customer revisit times. Note that the simulated revisit customer arrival times may be interspersed between the real data arrival times for the first time customers. We assume that the group size of a return customer group is identical to the group size of the same customer group during its previous visit. If the customer revisit time falls during the restaurant non-operating hours, the revisit time is adjusted to 12 noon of the following day.

The steps followed for each customer group in the full model simulation are illustrated in the flowchart (see Fig. 8). The flow is similar to the basic model except for the reneging probability and customer revisit time estimations. After a customer group arrives, a table category is assigned to the customer. Subsequently, the customer waits for a table from that table category to be available. Once the table is available, the model estimates the chance that the customer will renege as a result of the waiting time experienced at the restaurant (and the other predictors shown in Model 2 in Table 4). If the customer reneges, then the time until revisiting is estimated (using Model 5) and the customer waits until coming back at the estimated day. If the customer is seated, then the customer revisit time is estimated after the dining time and the customer waits until the estimated revisiting day occurs.

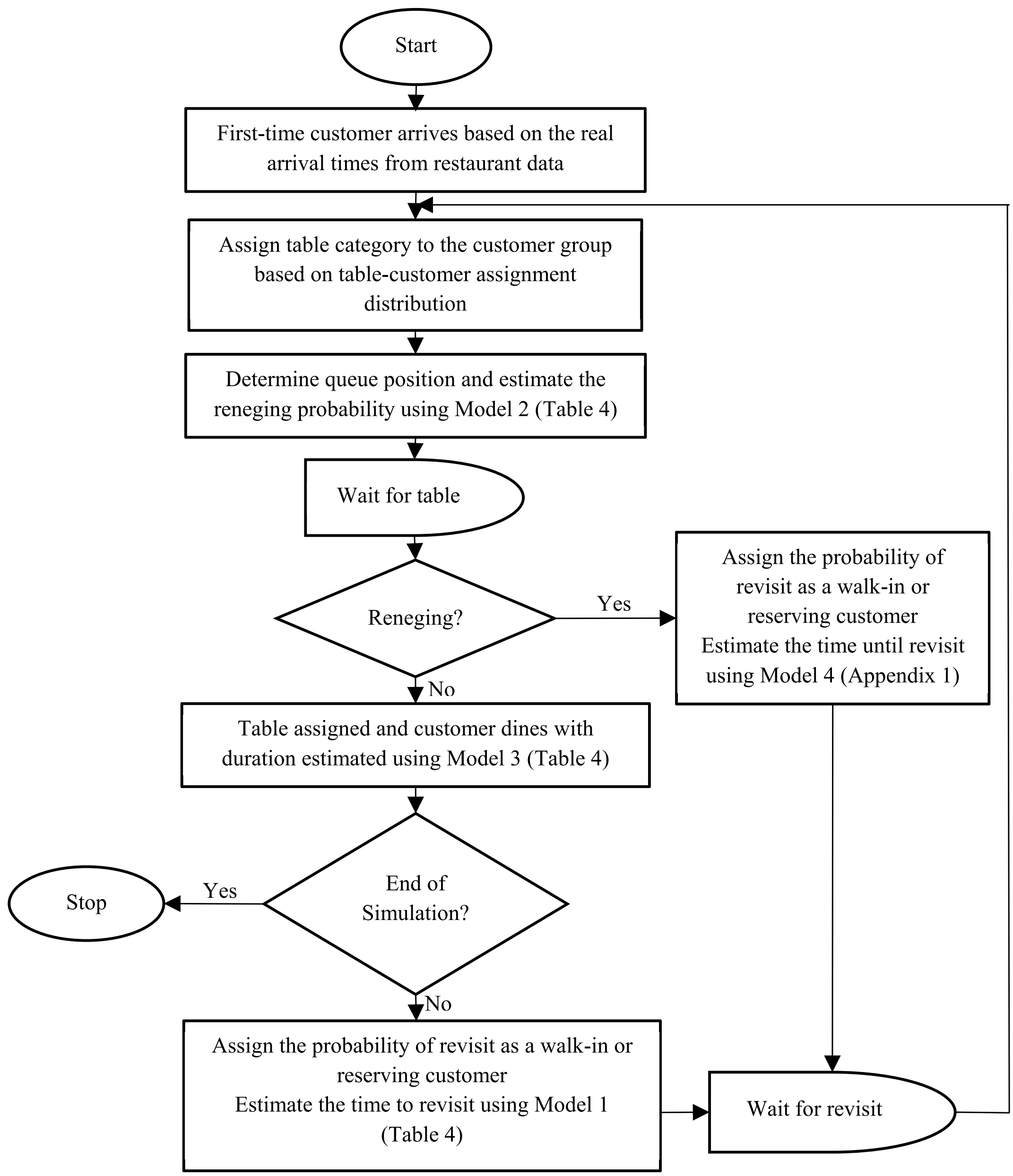

Fig. 8. Full model with steps followed for each customer group.

Basic model waiting time validation

We ran the basic simulation for a single run with a length of 274 days, using all real customer arrivals from February 1, 2016 up to and including October 31, 2016. At the beginning of each day we added a random wait from the distribution of waits faced by customers visiting right after the restaurant opens. This was done to account for the fact that the restaurant gradually prepares more tables. During the timeframe of the simulation, 73,159 customer groups (first timers and returning customers) visited the restaurant. The simulated waiting time distributions for customers visiting in various group sizes are compared with the actual waiting time distribution observed at the restaurant. Table 9 and Fig. 9 show that actual and simulated waiting times experienced are very similar.

Table 9
Descriptive statistics for the actual and simulated waiting times (in minutes)

Variable	N	Mean simulation	Mean actual	SD simulation	SD actual
Waiting time single diners	322	7.49	6.91	18.27	17.62
Waiting time groups of two	22,179	15.79	14.13	27.40	22.61
Waiting time groups of three	10,841	19.02	19.02	28.45	27.45

(continued on next page)

Table 9 (*continued*)

Variable	N	Mean simulation	Mean actual	SD simulation	SD actual
Waiting time groups of four	8236	19.02	21.22	29.96	30.65
Waiting time groups of five	3738	20.46	22.56	32.95	31.97
Waiting time groups of six	2450	27.48	24.17	34.23	34.23
Waiting time groups of seven	1155	28.55	24.96	56.46	33.44

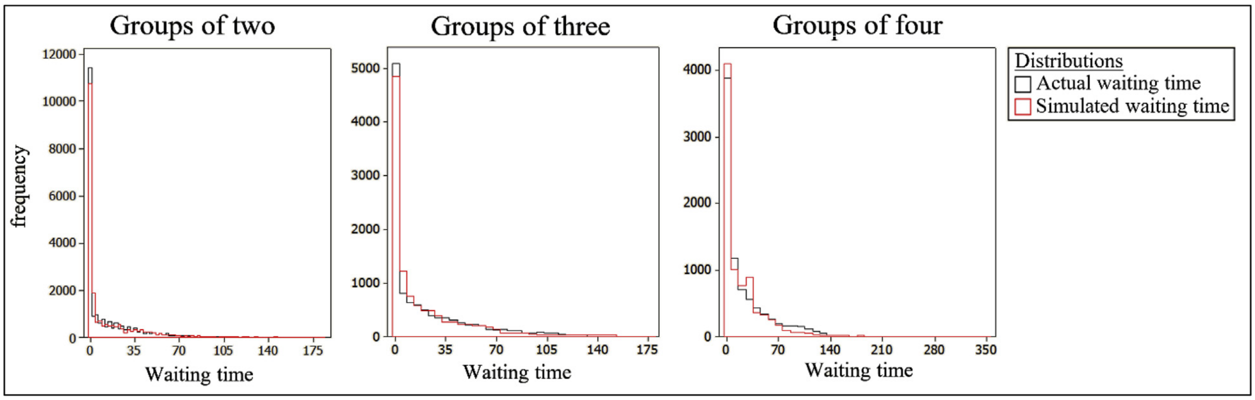

Fig. 9. Comparison of waiting time distributions between simulated and actual values for group sizes two, three, and four.

Full model validation of customer revisit distribution

In the full model, the customer arrivals were combined from two sources: real restaurant data for first time arrivals over the whole period of 12 months and simulation-based for returning customers. We first identified the visit time information for all customer groups that showed up for the first time during the 12 month time period, resulting in a list of 70,579 unique customer groups. We used the customer arrival time (weekend vs. weekday, evening vs. afternoon), customer group size, type of transaction (walk-in vs. reservation), and waiting time or queue position to estimate the chance the customer reneges and the time of a revisit.

Using our full simulation model, we validated the distribution of the customer groups that revisit with the revisits realized in the last three months. Table 10 shows the frequency distribution of the total number of customer groups returning to the restaurant and the number of reneging customers in the last three months. We chose the last three months of the data for out-of-sample, because the empirical models were built using the first 9 months data sample. For both statistics, we see from Table 6 that the multi-method approach (simulation combined with empirical models) provides solid out-of-sample estimates of the frequency distribution of customer visits for different group sizes.

Table 10
Total number of customer groups returning and reneging in the last three months

Group size	Return frequency		Renege frequency	
	Simulation	Actual	Simulation	Actual
1	55	77	14	10
2	4134	3899	1408	1416
3	1875	1765	695	661
4	1391	1382	524	490
5	530	553	216	205
6	343	382	131	138
7	133	122	48	46

To establish that our model is also delivering substantially better out-of-sample predictions than a model that does not incorporate dynamic effects, we employ an additional validation step. The dynamics in the simulation model are caused by the combination between the various empirical effects we identified: the impact of waiting time on reneging, on dining duration, and on time until revisiting. Removing the dynamic effects therefore means running a simulation in which the empirical equations are not based on waiting time. To do this, we remove waiting time as predictor from our empirical deterministic models and estimate the coefficients of the remaining predictors. These equations are subsequently used in the simulation model. Comparing the results (displayed in Tables 11 and 12) between the dynamic and non-dynamic simulation demonstrates that the dynamic simulation fits the actual data substantially better, especially for the most commonly occurring group sizes (2–5 persons). Not including waiting time as a predictor of reneging leads to a substantial overestimation of reneging and returning customers.

Table 11
Out of sample validation of the total number of customer groups reneging in the last three months, with and without dynamic effect of waiting time

Renege frequency					
Group size	Actual	Simulation	Deviation between simulation and actual	Non-dynamic simulation	Deviation between non-dynamic simulation and actual
1	10	14	+40.0%	28	+180.0%
2	1416	1408	−0.6%	1847	+30.4%
3	661	695	+5.1%	902	+36.5%
4	490	524	+6.9%	615	+25.5%
5	205	216	+5.4%	274	+33.7%
6	138	131	−5.1%	152	+10.1%
7	46	48	+4.3%	70	+52.2%

Table 12
Out of sample validation of the total number of customer groups returning in the last three months, with and without dynamic effect of waiting time

Return frequency					
Group size	Actual	Simulation	Deviation between simulation and actual	Non-dynamic simulation	Deviation between non-dynamic simulation and actual
1	77	55	−28.6%	66	−14.3%
2	3899	4134	+6.0%	4386	+12.5%
3	1765	1875	+6.2%	2015	+14.2%
4	1382	1391	+0.7%	1431	+3.5%
5	553	530	−4.2%	596	+7.8%
6	382	343	−10.2%	354	−7.3%
7	122	133	+9.0%	160	+31.1%

References

Afeche, P., 2013. Incentive-compatible revenue management in queueing systems: optimal strategic delay. Manuf. Serv. Oper. Manag. 15, 423–443.
Anand, K.S., Paç, M.F., Veeraraghavan, S., 2011. Quality–speed conundrum: trade-offs in customer-intensive services. Manag. Sci. 57, 40–56.
Antonides, G., Verhoef, P.C., Van Aalst, M., 2002. Consumer perception and evaluation of waiting time: a field experiment. J. Consum. Psychol. 12, 193–202.
Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear Mixed-effects Models Using Eigen and S4. R Package Version 1. pp. 1–23.
Batt, R.J., Terwiesch, C., 2015. Waiting patiently: an empirical study of queue abandonment in an emergency department. Manag. Sci. 61, 39–59.
Becker, G.S., 1965. A theory of the allocation of time. Econ. J. 493–517.
Bell, S.J., Auh, S., Smalley, K., 2005. Customer relationship dynamics: service quality and customer loyalty in the context of varying levels of customer expertise and switching costs. J. Acad. Market. Sci. 33, 169–183.
Bender, R., Augustin, T., Blettner, M., 2005. Generating survival times to simulate Cox proportional hazards models. Stat. Med. 24, 1713–1723.
Bielen, F., Demoulin, N., 2007. Waiting time influence on the satisfaction-loyalty relationship in services. Manag. Serv. Qual.: Int. J. 17, 174–193.
Chandon, P., Morwitz, V.G., Reinartz, W.J., 2005. Do intentions really predict behavior? Self-generated validity effects in survey research. J. Market. 69, 1–14.
Cox, T., 2005. Scientists Seek to Make Waiting in Line More Bearable [WWW Document]. News & Notes. http://www.npr.org/templates/story/story.php?storyId=4271009, Accessed date: 16 January 2017.
Crosby, L.A., Evans, K.R., Cowles, D., 1990. Relationship quality in services selling: an interpersonal influence perspective. J. Market. 68–81.
Dabholkar, P.A., 1990. How to improve perceived service quality by increasing customer participation. In: Proceedings of the 1990 Academy of Marketing Science (AMS) Annual Conference. Springer, pp. 483–487.
Dai, J.G., He, S., 2011. Queues in service systems: customer abandonment and diffusion approximations. In: Transforming Research into Action. INFORMS, pp. 36–59.
Davis, M.M., Heineke, J., 1998. How disconfirmation, perception and actual waiting times impact customer satisfaction. Int. J. Serv. Ind. Manag. 9, 64–73.
Davis, M.M., Maggard, M.J., 1990. An analysis of customer satisfaction with waiting times in a two-stage service process. J. Oper. Manag. 9, 324–334.
De Ruyter, K., Bloemer, J., 1999. Customer loyalty in extended service settings: the interaction between satisfaction, value attainment and positive mood. Int. J. Serv. Ind. Manag. 10, 320–336.
Debo, L.G., Parlour, C., Rajan, U., 2012. Signaling quality via queues. Manag. Sci. 58, 876–891.
Devaraj, S., Matta, K.F., Conlon, E., 2001. Product and service quality: the antecedents of customer loyalty in the automotive industry. Prod. Oper. Manag. 10, 424–439.
Dube, L., Renaghan, L.M., Miller, J.M., 1994. Measuring customer satisfaction for strategic management. Cornell Hotel Restaur. Adm. Q. 35, 39–47.
Ellinger, A.E., Daugherty, P.J., Plair, Q.J., 1999. Customer satisfaction and loyalty in supply chain: the role of communication. Transport. Res. E Logist. Transport. Rev. 35, 121–134.
Estelami, H., 2000. Competitive and procedural determinants of delight and disappointment in consumer complaint outcomes. J. Serv. Res. 2, 285–300.
Garnett, O., Mandelbaum, A., Reiman, M., 2002. Designing a call center with impatient customers. Manuf. Serv. Oper. Manag. 4, 208–227.
Gavirneni, S., Kulkarni, V.G., 2016. Self-selecting priority queues with burr distributed waiting costs. Prod. Oper. Manag. 25, 979–992.
Greene, W.H., 2012. In: Econometric Analysis, seventh ed. Prentice Hall, Upper Saddle River, NJ, USA.
Hair, J.F., Black, W., Babin, B., Anderson, R., 2010. In: Multivariate Data Analysis,

seventh ed. Prentice Hall, Upper Saddle River, NJ, USA.
Hallowell, R., 1996. The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. Int. J. Serv. Ind. Manag. 7, 27–42.
Harvey, J., 1998. Service quality: a tutorial. J. Oper. Manag. 16, 583–597.
He, Q.-C., Chen, Y.-J., 2018. Revenue-maximizing pricing and scheduling strategies in service systems with flexible customers. Oper. Res. Lett. 46, 134–137. https://doi.org/10.1016/j.orl.2017.11.012.
Hilbe, J.M., 2011. Negative Binomial Regression. Cambridge University Press, Cambridge, UK.
Hopp, W.J., Iravani, S.M.R., Yuen, G.Y., 2007. Operations systems with discretionary task completion. Manag. Sci. 53, 61–77.
Hosmer, D., Lemeshow, 1999. Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley-Interscience, New York, NY.
Houston, M.B., Bettencourt, L.A., Wenger, S., 1998. The relationship between waiting in a service queue and evaluations of service quality: a field theory perspective. Psychol. Market. 15, 735–753.
Hubeaux, S., Rufibach, K., 2014. SurvRegCensCov: Weibull Regression for a Right-censored Endpoint with a Censored Covariate. ArXiv Preprint arXiv:1402.0432.
Hwang, J., Gao, L., Jang, W., 2010. Joint demand and capacity management in a restaurant system. Eur. J. Oper. Res. 207, 465–472.
Ittig, P.T., 1994. Planning service capacity when demand is sensitive to delay. Decis. Sci. J. 25, 541–553.
Ittig, P.T., 2002. The real cost of making customers wait. Int. J. Serv. Ind. Manag. 13, 231–241.
Jabs, J., Devine, C.M., 2006. Time scarcity and food choices: an overview. Appetite 47, 196–204.
Jennings, O.B., Pender, J., 2016. Comparisons of ticket and standard queues. Queueing Syst. 84, 145–202.
Kc, D.S., Terwiesch, C., 2009. Impact of workload on service time and patient safety: an econometric analysis of hospital operations. Manag. Sci. 55, 1486–1498.
Kimes, S.E., Barrash, D.I., Alexander, J.E., 1999. Developing a restaurant revenue-management strategy. Cornell Hotel Restaur. Adm. Q. 40, 129–183.
Kimes, S.E., Robson, S.K.A., 2004. The impact of restaurant table characteristics on meal duration and spending. Cornell Hotel Restaur. Adm. Q. 45, 333–346.
Kimes, S.E., Thompson, G.M., 2004. Restaurant revenue management at Chevys: determining the best table mix. Decis. Sci. J. 35, 371–392.
Kimes, S.E., Wirtz, J., Noone, B.M., 2002. How long should dinner take? Measuring expected meal duration for restaurant revenue management. J. Revenue Pricing Manag. 1, 220–233.
Koo, M., Fishbach, A., 2010. A silver lining of standing in line: queuing increases value of products. J. Market. Res. 47, 713–724.
Koole, G., Mandelbaum, A., 2002. Queueing models of call centers: an introduction. Ann. Oper. Res. 113, 41–59.
Kostami, V., Ward, A.R., 2009. Managing service systems with an offline waiting option and customer abandonment. Manuf. Serv. Oper. Manag. 11, 644–656.
Kremer, M., Debo, L., 2015. Inferring quality from wait time. Manag. Sci. 62, 3023–3038.
Leclerc, F., Schmitt, B.H., 1999. The Value of Time in the Context of Waiting and Delays, Consumer Value: a Framework for Analysis and Research. Routledge, Abingdon, UK.
Lind, J.T., Mehlum, H., 2010. With or without U? the appropriate test for a U-Shaped relationship. Oxf. Bull. Econ. Stat. 72, 109–118.
Luo, W., Liberatore, M.J., Nydick, R.L., Chung, Q.B., Sloane, E., 2004. Impact of process change on customer perception of waiting time: a field study. Omega 32, 77–83.
Maglaras, C., Yao, J., Zeevi, A., 2017. Optimal price and delay differentiation in large-scale queueing systems. Manag. Sci. 64 (5), 2427–2444.
McFadden, D., 1973. Conditional Logit Analysis of Qualitative Choice Behavior.
Melnyk, S.A., Pagell, M., Jorae, G., Sharpe, A.S., 1995. Applying survival analysis to operations management: analyzing the differences in donor classes in the blood donation process. J. Oper. Manag. 13, 339–356.
Munda, M., Rotolo, F., Legrand, C., 2012. Parfm: parametric frailty models in R. J. Stat. Software 51, 1–20.
Munichor, N., Rafaeli, A., 2007. Numbers or apologies? Customer reactions to telephone waiting time fillers. J. Appl. Psychol. 92, 511.
Nie, W., 2000. Waiting: integrating social and psychological perspectives in operations management. Omega 28, 611–629.
Osuna, E.E., 1985. The psychological cost of waiting. J. Math. Psychol. 29, 82–105.
Payne, A., Frow, P., 2005. A strategic framework for customer relationship management. J. Market. 69, 167–176.
R Core Team, 2017. R: a Language and Environment for Statistical Computing.
Rafaeli, A., Barron, G., Haber, K., 2002. The effects of queue structure on attitudes. J. Serv. Res. 5, 125–139.
Roy, D., Bandyopadhyay, A., Banerjee, P., 2016. A nested semi-open queuing network model for analyzing dine-in restaurant performance. Comput. Oper. Res. 65, 29–41.
Shafer, S.M., Smunt, T.L., 2004. Empirical simulation studies in operations management: context, trends, and research opportunities. J. Oper. Manag. 22, 345–354.
Shapiro, C., Varian, H.R., 2013. Information Rules: a Strategic Guide to the Network Economy. Harvard Business Press.
Shunko, M., Niederhoff, J., Rosokha, Y., 2017. Humans are not machines: the behavioral impact of queueing design on service time. Manag. Sci. http://dx.doi.org/10.1287/mnsc.2016.2610.
Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., Bolker, B., 2015. glmmADMB.
Srivastava, R.K., Shervani, T.A., Fahey, L., 1998. Market-based assets and shareholder value: a framework for analysis. J. Market. 2–18.
Stank, T.P., Goldsby, T.J., Vickery, S.K., 1999. Effect of service supplier performance on satisfaction and loyalty of store managers in the fast food industry. J. Oper. Manag. 17, 429–447.
Stone, A., 2012. Why Waiting Is Torture [WWW Document]. The New York Times. http://www.nytimes.com/2012/08/19/opinion/sunday/why-waiting-in-line-is-torture.html.
Tan, T.F., Netessine, S., 2014. When does the devil make work? An empirical study of the impact of workload on worker productivity. Manag. Sci. 60, 1574–1593.
Taylor, S., 1994. Waiting for service: the relationship between delays and evaluations of service. J. Market. 56–69.
Therneau, T.M., Lumley, T., 2017. Package “survival.” R Package Version 2–41.
Thompson, D.A., Yarnold, P.R., Williams, D.R., Adams, S.L., 1996. Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. Ann. Emerg. Med. 28, 657–665.
Thompson, G.M., 2009. (Mythical) revenue benefits of reducing dining duration in restaurants. Cornell Hospitality Quarterly 50, 96–112.
Tjur, T., 2009. Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. Am. Statistician 63, 366–372.
Umesh, U., Pettit, K., Bozman, C.S., 1989. Shopping model of the time-sensitive consumer. Decis. Sci. J. 20, 715–729.
Veeraraghavan, S., Debo, L., 2009. Joining longer queues: information externalities in queue choice. Manuf. Serv. Oper. Manag. 11, 543–562.
Venables, W.N., Ripley, B.D., 2002. In: Modern Applied Statistics with S, Fourth. ed. Springer, New York, NY.
Wang, J., Zhou, Y.-P., 2017. Impact of queue configuration on service time: evidence from a supermarket. Manag. Sci.
Wu, C.A., Bassamboo, A., Perry, O., 2018. Service systems with dependent service and patience times. Manag. Sci.
Yu, Y., Benjaafar, S., Gerchak, Y., 2015. Capacity sharing and cost allocation among independent firms with congestion. Prod. Oper. Manag. 24, 1285–1310.
Zauberman, G., Kim, B.K., Malkoc, S.A., Bettman, J.R., 2009. Discounting time and time discounting: subjective time perception and intertemporal preferences. J. Market. Res. 46, 543–556.
Zeithaml, V.A., 1988. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J. Market. 2–22.
Zhang, Z., 2016. Parametric regression model for survival data: Weibull regression model as an example. Ann. Transl. Med. 4.