CrossMark

Evaluating the performance of a dial-a-ride service using simulation

Carl H. Häll · Jan T. Lundgren · Stefan Voß

Abstract In this paper we study the effects of changes to a paratransit system operated as a dynamic dial-a-ride service. The scheduling of requests is limited by time window constraints and maximum ride time constraints; costs for customer discomfort in form of waiting time and excess ride time are considered in the planning of the service. The parameters defining these constraints and costs are evaluated based on simulations of a real-world scenario. Several different criteria, both regarding customer level of service and operational costs, are used to evaluate the simulation results. By showing which parameters have a large impact on the different evaluation criteria, we can give guidelines to operators of public transport of how to design their dial-a-ride services. Numerical results are presented and from these results conclusions are drawn regarding which parameters are most important in dynamic dial-a-ride services.

Keywords Dial-a-ride · Paratransit · Simulation

1 Introduction

Despite efforts of increasing public transport services availability to a general public, many people are unable to use the normal public transport services due to physical or mental impairment. It is therefore necessary to introduce some form of transportation service that is specially designed for these customers. This type of

C. H. Häll (✉) · J. T. Lundgren
Department of Science and Technology (ITN), Linköping University, 60174 Norrköping, Sweden
e-mail: carl.henrik.hall@liu.se

S. Voß
Institute of Information Systems, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany

🕭 Springer

service is called paratransit. It is often operated as a demand responsive service and is a complement to the regular public transport to provide a transport system that satisfies the mobility needs of the whole population.

One common form of paratransit is to operate it as a dial-a-ride service, where customers call in requests to a call center and the transportation is carried out from door to door. A service of this kind can be designed in many different ways, depending on the level of service the operator wishes to provide to the customers. One key issue is how to balance the service level against the operating cost of the service. Dial-a-ride is an expensive form of transportation service, and in order to develop a cost-efficient service, it is important to know how different system designs affect both the level of service for the customers and the cost for the operator.

When operating a dial-a-ride service, routes and schedules are to be designed for a fleet of vehicles, given a set of travel requests. The set of requests can include individual passengers or groups of passengers, with specified origins and destinations and a requested time for when the journey is to take place. The schedules have to satisfy passengers' requirements in terms of user inconvenience, e.g. defined by waiting times, maximum allowed ride times or deviations from desired departure and arrival times. The overall goal is to minimize a combination of user inconvenience and operating costs. This problem is known as the dial-a-ride problem and is a special case of the pick-up and delivery problem with time windows. For more information about the dial-a-ride problem, we refer to the surveys of Cordeau and Laporte (2007) and Parragh et al. (2008).

The aim of our work is to evaluate, by simulation, which effects variations of different parameters have on travel time and other service indicators as well as cost related indicators such as the number of vehicles required, total travel distance of the vehicles etc. By showing which parameters have a large impact on the efficiency and performance of the service, we can give guidelines to public transport operators about how to design their dial-a-ride services. Operators have for some years requested general knowledge about how the parameter values affect their services. Therefore, the purpose is not to find exact values for the specific instance used, but rather to show which parameters are likely to largely effect the service, and in which way.

We use our modeling system DARS, described in Häll et al. (2012), to evaluate the effects of these parameters by simulation. For the simulations we use real-world data from a dynamic dial-a-ride service operating in Sweden, i.e., the set of requests and the way the service is planned and operated is coming from an existing service. The concept of dynamic dial-a-ride service means that not all requests are known prior to the start of the planning, but increases as requests are called in.

Simulation has for many years been used to evaluate the effects of changes in dial-a-ride services. For example, Wilson et al. (1969) use simulation to evaluate how various planning algorithms affect the behavior of the service. Bailey and Clark (1987) study the interaction between demand, service rate and policy alternatives for a taxi service. In Fu (2002) simulation is used to evaluate what effects automatic vehicle location systems can have on a dial-a-ride service. In Shinoda et al. (2004), immediate-request dial-a-ride systems and fixed route systems are compared

through simulation. The aim is to compare the usability and profitability of the dial-a-ride systems to that of the fixed route systems. Quadrifoglio et al. (2008) use simulation to study time window settings and zoning vs. no-zoning strategies in a dial-a-ride service. They study how such changes affect the total trip time, deadhead miles and the fleet size. Deflorio (2011) presents a method to generate trip requests that has been proposed to analyze the performance of demand responsive services in situations/areas with low and sparse demand.

Simulation has also been used to study flexible forms of public transport services, such as a mobility allowance shuttle transit service, where a vehicle is allowed to deviate from its fixed path to pick up, or drop off, customers within a service area. Such a service is studied, by simulation, in Quadrifoglio and Dessouky (2008) to show a sensitivity analysis of how the performance of the system is depending on the shape of the service area.

Previous research has not considered the effects of all the parameters we evaluate in this paper for one and the same real-world service. Nevertheless, in some ways, our evaluations and methodology resembles that of Quadrifoglio et al. (2008). Therefore, we will later also discuss some of our findings in comparison to those of that paper.

Section 2 describes how the paratransit service we consider in this study is operated and also presents the input regarding geography, requests and vehicles that is considered in our scenario. In Sect. 3, we describe for which parameters variation effects are evaluated. Also described is how the planning of the considered service is done and what cost parameters are used in the planning. Section 4 describes which evaluation criteria are used and the results of all performed simulations. A further discussion of some of these results is then provided in Sect. 5. Finally, in Sect. 6, the conclusions of this study are presented.

2 Test scenario Norrköping

Operators of paratransit services, operated as dynamic dial-a-ride systems, often do not know what effect different choices of parameter settings have on the provided solutions from the planning system they use. For this reason it is very important to find out for realistic instances, how changes to the most important parameters affect the solution. This is also the reason why we base our evaluation on sensitivity analysis on a few important parameters. For the analysis, we use real-world input regarding the network, vehicle fleet, requests as well as the standard parameter values we use in our base scenario to which we evaluate all changes to the parameter settings.

Our base scenario is a description of how input from the real-world paratransit system in the city of Norrköping, Sweden, would perform with the planning method and parameter values that are often used in planning of paratransit in Sweden.

In this service, customers can call in requests at any time, requesting transportation from any location to any other location (within the service area that the municipality is responsible for). The customer also gives a requested pick-

up time and information whether the customer uses wheelchair or if he, or she, requires extra legroom in the vehicle (further described in Sect. 2.2).

While the customer is still on the phone, the planning system calculates the best insertion of the new request into the existing schedule (all vehicle routes). The insertion must be made in such a way that the customer is picked up within a given time window (TW1) centered around the requested pick-up time, and guaranteeing that a maximum ride time is not exceeded. A replanning of the pick-up time can later be made within a second time window, TW2. Requests may be planned to be served jointly by a vehicle (shared vehicle service).

This is the most common way to operate this type of service in Sweden. The service will be further described in Sect. 3.2, where the parameter values used in the base scenario of our simulations are described. The remainder of this section describes the input, in form of the geography, requests and vehicles that are considered in the test scenario.

2.1 Service area

The service covers a geographical area of approximately 630 square miles. The area is divided into 1,064 zones and every address in the service area belongs to one such zone.

Between each pair of zones there is a precalculated, estimated travel time. These estimated travel times are used as input in the planning process. As the estimated travel time between two addresses is based on zones, all estimated travel times are therefore independent of which exact addresses, in the zones, are really considered. Travel times between two locations in the same zone are also considered in the same way, i.e., an estimated travel time exists but no regard is taken to whether they are next door or as far apart as possible within the zone.
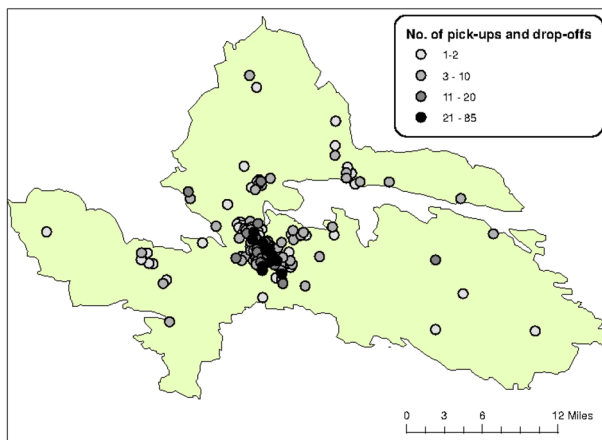
Fig. 1 Description of the quantity of pick-ups and drop-offs occurring in the service area

All travel times can be expressed by some value in a $1,064 \times 1,064$ matrix. Several such matrices can be used if needed, e.g. for representing various times of the day, various weather conditions etc. In our case we use the same matrix for the whole day. The way that the estimated travel times are defined is of course a simplification, but as the simplification is made in the real-world planning system we will use the same assumptions.

2.2 Requests

The set of requests we consider in this evaluation is taken from one day's operation of the Norrköping paratransit service. The set includes 606 requests. The pick-ups and drop-offs of these requests are distributed over 164 of the total 1,064 zones. How the requests are distributed is shown in Fig. 1, and a more detailed visualization of the central parts of Norrköping is given in Fig. 2.

Three different types of customers are considered; ambulatory customers, customers who require extra legroom (known as front-seat riders, since they can be served by a regular taxi vehicle if they are seated in the front seat) and customers in a wheelchair. 25 % of the requests involve transportation of a customer using wheelchair. 76 out of the 606 requests include an additional passenger, giving a total of 682 persons to be transported. Out of these are 70 % ambulatory customers, 8 % are front-seat riders and 22 % use a wheelchair.

58 % of the requests are called in at least the day before they are to be served, some several weeks in advance. The remaining requests are called in according to the distribution presented in Fig. 3. In that figure the distribution of requested pick-up times is also described. The requested pick-up times in the figure have been discretized hour by hour.

The time it would take to transport a customer directly from pick-up to drop-off is referred to as the direct ride time. The distribution of the direct ride time is given
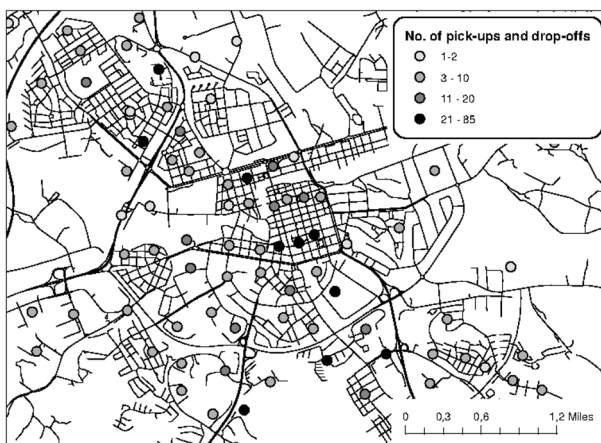
Fig. 2 Description of the quantity of pick-ups and drop-offs occurring in the central parts of the city of Norrköping
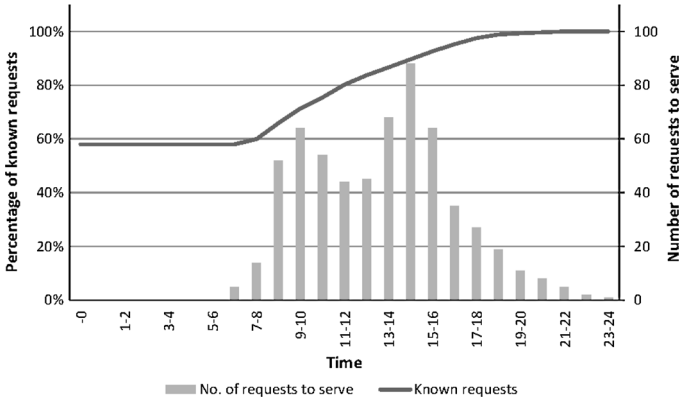
Fig. 3 Description of call-in times and requested pick-up times

in Fig. 4, where the times are discretized in 2-min intervals. The average direct ride time of the requests is 10.6 min and 90 % of the requests have a direct ride time that is less than 20 min.

The presented set of requests represents a typical weekday (Monday–Thursday). A large part of the requests, about one third, represents trips to and from places of work and daily activity centers. This means that there is a good base of requests that will be repeated throughout the weekdays. On Fridays, the requests normally have another distribution over time and on Saturdays and Sundays the number of requests also differs from that of weekdays. On major holidays, such as Christmas and New Year, the number of requests is quite much higher than on a regular weekday and the average distance of the requested trips is longer.

However, for weekdays, our set of requests is representative. Still, for some of the evaluation criteria that will be presented in Sect. 4, the fact that only one day of requests is used will affect the possibility to interpret the results. This will be discussed after the description of the criteria.
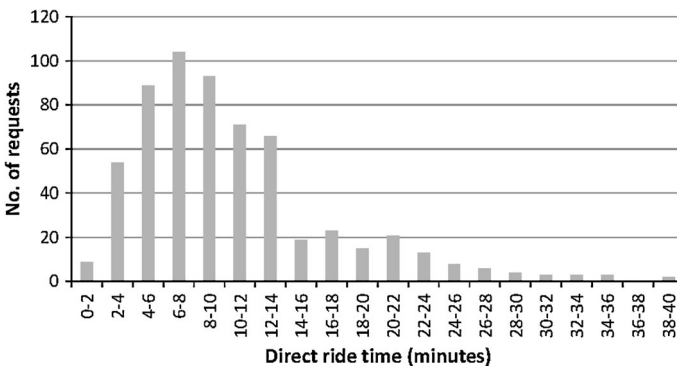
Fig. 4 Illustration of the direct ride time distribution

2.3 Vehicles

There are two different types of vehicles available, ordinary taxi vehicles and vans equipped to transport persons seated in wheelchairs. Each type of vehicle has a limited capacity regarding each type of customer (described in the previous section). A front-seat rider can, however, be seated in a regular seat in a van, but only in the front-seat of a taxi vehicle. The capacities are presented in Table 1. An ambulatory customer can use a seat that is spacious enough for a front-seat rider, but a front-seat rider cannot use a regular seat for an ambulatory customer.

The number of vehicles, both taxi vehicles and vans, can be regarded as unlimited. This is the case since the paratransit operator has agreements with several taxi companies to use their vehicles. If the vehicles are not needed in the paratransit service, they are used in the regular taxi service.

In some practical instances, the number of vans is a limited resource (i.e. on some specific days of the year when the demand is at its peak). However, in our case both vehicle types are unlimited resources. Nevertheless, each vehicle can be operated at maximum 4 h in one shift and each new vehicle that is taken into operation (each new shift that is started) has a start-up cost related to it.

3 Simulation description

In this section we describe the parameters to be evaluated in our simulations. We also describe how the planning of the considered service is made (in the real-world and in our simulation model), and also which parameter values are used in the base scenario that we compare our other simulation results to.

3.1 Simulation environment

To perform the sensitivity analysis, i.e. to evaluate the effects of the various parameter settings, we have used the modeling system DARS described in Häll et al. (2012). Since we want to perform a sensitivity analysis on an existing service, we have not developed any planning algorithm in this study. Instead, we merely use DARS as a platform to mimic the characteristics from the real-world planning system used for the operational planning of the paratransit service. One possibility would of course have been to use the existing real-world planning system, but we

Table 1 Vehicle capacities

Vehicle type	Number of ambulatory customers	Number of front-seat riders	Number of wheelchair customers
Taxi	3	1	0
Van	0	3	3

have found it easier to control both the input data and the output statistics using DARS.

DARS consists of four modules; Input manager, Simulation control, Request manager and Output generator. All information about the scenario (regarding costs, requests, vehicles etc.) is loaded to the Input manager prior to simulation. When the simulation is started, the set of requests is sent to the Simulation control and when the simulation time equals the call-in time of a specific request, that request is passed on to the Request manager. In the Request manager, the request is inserted into the schedule in the position that gives the lowest increase to the objective function value. This is done while ensuring that the constraints regarding TW1 and maximum ride time are not violated for that request, and that constraints regarding TW2 and maximum ride time (as described in Sect. 2) are fulfilled for all scheduled requests.

The main purpose of the Simulation control is to keep track of time and make sure that various actions are notified to other parts of DARS at the right time. The Output manager summarizes the statistics and generates the output of each simulation run, while the role of the Request manager is to handle the schedule; i.e. to insert requests and replan existing schedules.

The planning system is constantly searching for improvements to existing schedules. Such replanning is done by randomly removing one request at a time from the schedule and search for better positions to reinsert it into, without violating constraints regarding TW2 and maximum ride time for any scheduled request. Other ways to perform the replanning by use of various ruin and recreate methods have been studied and shown to give very positive results. That study is presented in Häll and Peterson (2013).

In both the insertion of new requests and replanning of existing requests, the goal is to minimize the objective function based on the calculation of the total cost function z, defined as:

$$z = \sum_{k=1}^{K}(FC_k + TC_k) + \sum_{i=1}^{I}(WTC_i + ETC_i)$$

where, FC_k is the fixed cost associated with the use of vehicle k, TC_k is the time related cost of vehicle k, WTC_i is the waiting time cost of request i, ETC_i is the excess ride time cost of request i, K is the number of vehicles, and I is the number of requests

How these costs are set in our base scenario, is described in the next section (Sect. 3.2). For a more detailed description of DARS, including construction of schedules and exact formulations of constraints for TW1 and TW2, see Häll et al. (2012).

3.2 Base scenario settings

The settings of the base scenario very well represent how a large part of the existing Swedish dial-a-ride services are operated. The fixed cost of taking a new vehicle into operation is of course a monetary value itself, and is given in the Swedish

currency (SEK). It is SEK 25 for a taxi and SEK 30 for a van. The time related cost is the time, in hours, that a vehicle is used multiplied by a cost parameter (i.e. a linear function). This parameter is for taxi vehicles SEK 270 and for vans SEK 300. The excess ride time is defined as the difference between the ride time (pick-up time to drop-off time) and the direct ride time. Both the excess ride time cost and the waiting time cost are defined by non-linear functions as described in Fig. 5. Both costs are given in SEK.

Regarding the customer related constraints (service constraints) the parameters in the base scenario are set in the following way. TW1 is set to 30 min, giving that a new request must be inserted ±15 min of the requested pick-up time. TW2 is set to 10 min, giving that it is acceptable to replan requests as long as all requests are picked up within 10 min from their promised pick-up time. The maximum ride time of request i (MRT_i) is depending on the direct ride time of that request (DRT_i) set to:

$$MRT_i = \begin{cases} \alpha \times DRT_i & \text{if } DRT_i \geq 15 \text{ min} \\ \alpha \times DRT_i + 15 \text{ min} & \text{if } DRT_i < 15 \text{ min} \end{cases} \tag{1}$$

where α is set to 2.

In our model, we do not consider any stochastic events (such as no-shows, vehicle break-downs etc.). However, some stochasticity exists in the planning system itself, due to the fact that in the real-world planning system, there is some randomness regarding which requests are chosen in the replanning phase and regarding which vehicles are chosen in some situations. For example, let us assume that there are two vehicles waiting at the same location, and that both are free for receiving new assignments (no more requests are planned at them for the moment). Further, assume that one of the vehicles has a maximum of 1 h left on the shift and that the second vehicle has at most 3 h left on its shift. When a new request that should be planned on one of the two vehicles is called in, the objective function value is not affected by which vehicle is chosen. However, later on the choice of vehicle can affect the objective function value if any additional requests are called in that would be suitable to plan on the same route if the vehicle with more time left on its shift would have been chosen.
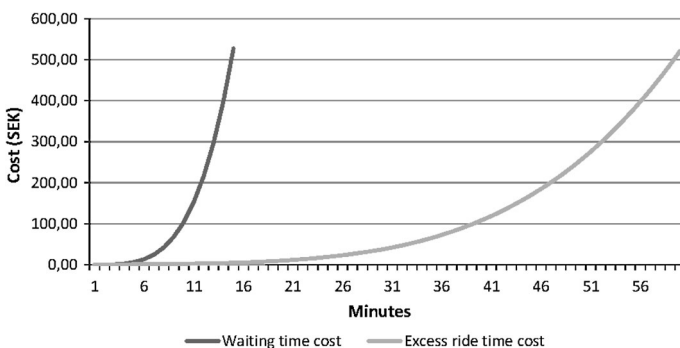
Fig. 5 Description of waiting time costs and excess ride time costs

So, even though we do not consider stochastic events in our model, the cost-structure (and planning method) itself determines that several repetitions of each parameter setting are necessary. For this reason we have chosen to perform five simulations of each parameter setting and compare the average values.

3.3 Evaluated parameters

There are several different cost parameters and service parameters of which it is interesting to study their effects. We study the following parameters: time window sizes (TW1 and TW2), maximum allowed ride time, waiting time costs and excess ride time costs. Time window sizes and maximum allowed ride time, are parameters that describe the level of service that is provided. Waiting time costs and excess ride time costs describe how user inconvenience is weighted (in the objective function) in comparison to the operating costs.

The evaluation of each parameter is based on several simulations. The simulations are done to identify not only which parameters have a larger effect on the service, but also to be able to identify how the effects of a certain parameter are depending on the parameter value.

4 Results

As described in Sect. 3.3, five parameters are evaluated. For each setting of a specific parameter value, five simulations are made. The average values of the five simulation runs for a specific value are then evaluated against the average values of the base scenario (for which also five simulations are run). The results are evaluated from nine different criteria. These evaluation criteria are:

- No. of vehicles (no. of started shifts)
- Total objective function value
- Total vehicle ride time
- Total vehicle ride time with passengers
- Total vehicle ride time without passengers
- Total request excess ride time
- Worst excess ride time
- Total deviation from promised pick-up time
- Worst deviation from promised pick-up time.

To find good recommendations to service providers, one would wish to perform more tests, on several instances (set of requests representing several days), test more values for each parameter and also make more simulation runs for each parameter value. However, based on the amount of time each simulation run takes, the study has to be limited to cover a reasonable amount of simulation runs. For this reason it is very important to provide good analysis of the results, and therefore t tests are performed to provide a p value for all evaluation criteria. The p value represents the probability that the difference from the base scenario is due to sampling error (due

to insufficient number of simulation runs). Therefore, a p value close to zero gives high significance to the result. For each test, important and significant (95 %) effects will be commented. Evaluation criteria that show no, or very minor, effects will be uncommented.

As mentioned before, the fact that requests from only one day are used affects the interpretation of the results. Since our chosen day's requests are very well representable for any weekday, and since we perform five simulations for each parameter value that is evaluated, the values of most criteria can be thought of as generalizable. This means that for any instance that uses the same parameter values as in the base scenario (presented in Sect. 4.1), similar effects will be seen when one of the parameter values is changed. However, for the criteria; *Worst excess ride time* and *Worst deviation from promised pick-up time*, this is not necessarily true. These values of these criteria can be very specific for each instance, and should therefore, even if the p value indicates a high significance to the change, not be generalized to be valid for any instance.

4.1 Base scenario

The average values of the base scenario simulations are given in Table 2. All time-related values are given in seconds. The table shows that almost half of the vehicle drive time is without any passengers. It can also be seen from the total request excess ride time that the average excess ride time of the 606 requests is about 1.4 min, and that the average deviation from promised pick-up time is only about 17 s.

This means that almost all requests are served very close to the promised pick-up times that are given to the customers, and it indicates that very few requests are served by a shared vehicle. These results are due to the fact that the cost of customer waiting time is set so high (as earlier described in Fig. 5).

In the following sections (Sects. 4.2–4.6) the results of the simulations of the various parameter settings will be presented. For those results it will be presented how much (in percentage) they deviate from the average values of this base scenario (i.e. from the results presented in Table 2). Before doing that we must, however, first take a look at how much the five simulation runs of the base scenario alter

Table 2 Results of base scenario simulations

Evaluation criterion	Value
No. of vehicles (no. of started shifts)	89.6
Total objective function value	75,797.4
Total vehicle ride time	723,062.0
Total vehicle ride time with passengers	385,711.6
Total vehicle ride time without passengers	337,349.8
Total request excess ride time	51,496.4
Worst excess ride time	1,473.0
Total deviation from promised pick-up time	10,345.8
Worst deviation from promised pick-up time	467.2

among themselves. This is presented in Fig. 6, where the average value (of the five simulations) of an evaluation criterion is represented by the 100 % level, and all individual results are indicated by a marker. It can be seen that the vehicle-related criteria (no. of vehicles, objective function value and travel time) have smaller differences between each simulation run than the service-related criteria (excess ride time and deviation from promised pick-up time). The largest deviation of the number of vehicles used is almost 4 % while for the other vehicle-related criteria the largest deviations are within approximately 2 % from their average values. This shows that for these criteria, the results of the five simulation runs give approximately the same values. However, for the service-related criteria the deviations vary up to about 20 % (worst excess ride time). This is why the results of the t tests are important to consider when the effects on the evaluation criteria are studied.

4.2 Changes to TW1

Regarding TW1, two different values (other than that used in the base scenario) are tested. In the first test the value is reduced from 30 to 15 min, and in the second test the value is increased to 60 min. This means that 50 and 200 % of the original value are evaluated. The results of the tests are presented in Table 3.

Not surprisingly, the total objective function value is increased when TW1 is reduced, and somewhat decreased when TW1 is increased. What is a bit more surprising is the large effect that can be seen on the number of vehicles that are used. When TW1 is reduced to 15 min the number of vehicles used is increased by 16 %. The reduced possibility to create good schedules makes it more beneficial to "let vehicles go" instead of having them waiting for a next request to serve, and then require new vehicles again. More but shorter routes are created.
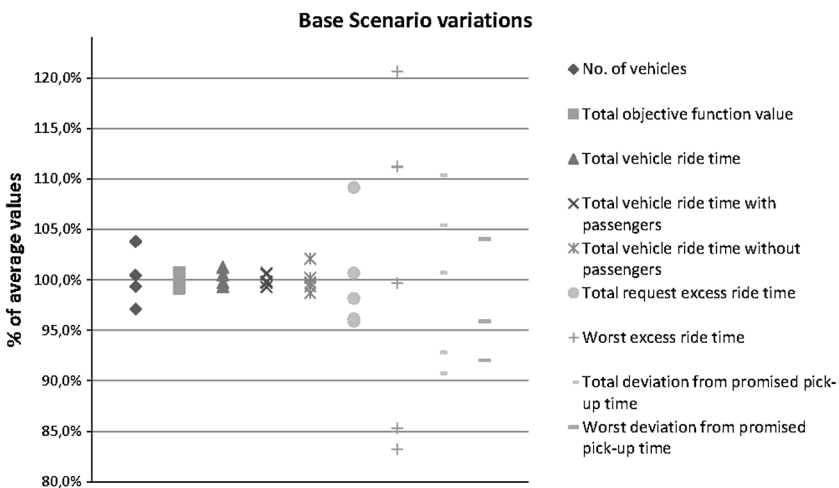

Fig. 6 Variation of the evaluation criteria results for the five simulations of the base scenario

Table 3 Results of TW1 simulations

Evaluation criterion	TW1 = 15 (%)	p value	TW1 = 60 (%)	p value
No. of vehicles (no. of started shifts)	16.3	0.00	−33.5	0.00
Total objective function value	4.3	0.00	−1.2	0.01
Total vehicle ride time	1.9	0.00	0.0	0.93
Total vehicle ride time with passengers	−1.3	0.01	−1.0	0.01
Total vehicle ride time without passengers	5.5	0.00	1.3	0.09
Total request excess ride time	−1.6	0.65	−12.4	0.01
Worst excess ride time	−2.9	0.73	−9.6	0.25
Total deviation from promised pick-up time	26.6	0.00	−53.9	0.00
Worst deviation from promised pick-up time	2.5	0.77	−15.5	0.00

When TW1 is increased to 200 % the number of vehicles used is reduced by 33 %. This would certainly be appreciated by the taxi-companies, since they will have information further in advance regarding what vehicles that will be used in the paratransit service.

From the results we can also see that the ride time without passengers is increased when TW1 is decreased to 15 min. This means that worse possibilities to create good schedules have increased the dead-head miles. This has also led to an increase of the total deviation from promised pick-up time by 27 %, meaning that more customers have to wait before the vehicle arrives, and the opposite occurs when TW1 is increased. However, note that the total deviation from promised pick-up time (the total waiting time of the customers) was very low in the base scenario.

The effects to all the criteria discussed in this section show a significant change (95 %, since p values ≤ 0.05).

4.3 Changes to TW2

From the results of the tests regarding TW2, presented in Table 4, there are only a few things worth mentioning. TW2 ensures that a pick-up is made within TW2-min

Table 4 Results of TW2 simulations

Evaluation criterion	TW2 = 5 (%)	p value	TW2 = 20 (%)	p value
No. of vehicles (no. of started shifts)	2.0	0.20	0.7	0.67
Total objective function value	0.8	0.04	0.9	0.03
Total vehicle ride time	1.0	0.06	1.1	0.06
Total vehicle ride time with passengers	0.0	0.93	0.7	0.07
Total vehicle ride time without passengers	2.1	0.02	1.6	0.10
Total request excess ride time	2.6	0.39	−1.5	0.68
Worst excess ride time	−15.6	0.10	−3.9	0.62
Total deviation from promised pick-up time	−22.8	0.00	−9.5	0.12
Worst deviation from promised pick-up time	−41.4	0.00	7.1	0.52

from the promised pick-up time. In the base scenario, the worst deviation (for one single request) is 7.8 min (on average over the 5 simulation runs). So, when TW2 is decreased to 5 min, from the 10 min that were used in the base scenario, the worst deviation from promised pick-up time is of course decreased. This is due to the fact that the TW2 constraint now states that the deviation from promised pick-up time may not be more than 5 min. So, the change is definitely significant, having p values very close to 0. Also the total deviation from the promised pick-up time is significantly reduced.

However, when TW2 is increased to 20 min we cannot see any significant change for these criteria. The way the waiting time cost function is defined makes it unprofitable to use the extension of TW2. By this reason, with TW2 set to 20 min, there is not much difference to any of the evaluation criteria compared to the base scenario, as can be seen by the overall high p values.

4.4 Changes to maximum ride time

In the tests regarding the maximum ride time, the factor α of Eq. 1 is reduced to 1.5 (from 2 in the base scenario) and increased to 2.5. The results presented in Table 5 show that with a reduced value of α, the total vehicle ride time without passengers is increased a little due to the fact that the new, harder constraints regarding maximum ride time gives worse possibilities to create good routes. The total excess ride time is also affected, it is decreased with 11 %. But here we should remember that this value was very low in the base scenario. The results also show small, but significant (p values ≤ 0.05), changes to total vehicle ride time and total vehicle ride time with passengers.

When α is increased, there is a significant change to the worst excess ride time, but as discussed earlier, the change to this criterion can be depending on the specific instance (the set of requests used). Therefore, the results regarding this criterion should not be generalized.

All other p values for $\alpha = 2.5$ show us that no other criteria have had any significant change. So, most likely, increasing α to 2.5 would have no effect on the behavior of the service.

Table 5 Results of maximum ride time simulations

Evaluation criterion	$\alpha = 1.5$ (%)	p value	$\alpha = 2.5$ (%)	p value
No. of vehicles (no. of started shifts)	−1.3	0.31	−0.4	0.74
Total objective function value	0.6	0.09	0.7	0.08
Total vehicle ride time	1.5	0.01	0.1	0.81
Total vehicle ride time with passengers	−0.7	0.05	−0.1	0.71
Total vehicle ride time without passengers	3.9	0.00	0.4	0.66
Total request excess ride time	−10.7	0.01	−0.5	0.89
Worst excess ride time	−16.8	0.08	23.2	0.03
Total deviation from promised pick-up time	3.4	0.53	−2.8	0.67
Worst deviation from promised pick-up time	4.0	0.19	3.7	0.30

Table 6 Results of waiting time cost simulations

Evaluation criterion	WTC = 10 % (%)	p value	WTC = 50 % (%)	p value
No. of vehicles (no. of started shifts)	−6.5	0.00	−6.9	0.00
Total objective function value	−0.9	0.04	0.7	0.08
Total vehicle ride time	−2.0	0.00	0.8	0.11
Total vehicle ride time with passengers	−0.3	0.34	0.1	0.79
Total vehicle ride time without passengers	−3.9	0.00	1.6	0.05
Total request excess ride time	46.0	0.00	3.8	0.28
Worst excess ride time	27.4	0.02	−0.6	0.94
Total deviation from promised pick-up time	115.4	0.00	16.7	0.01
Worst deviation from promised pick-up time	16.1	0.00	4.9	0.13

4.5 Changes to waiting time costs

When reducing the waiting time cost to half ($WTC = 50\%$), the only major and significant changes occur to the number of vehicles and the total deviation from promised pick-up time, as can be seen in Table 6. Longer waiting times are planned for, and therefore fewer vehicles are needed. All other criteria have minor or insignificant changes.

When the cost function is reduced to one tenth ($WTC = 10\%$), the values of all criteria but one are changed significantly (having p values ≤ 0.05). The total deviation from promised pick-up time is considerably increased, which indicates that more requests are being rescheduled using the flexibility of TW2. The worst deviation from promised pick-up time is also increased, now approaching 10 min, i.e. the largest deviation that is allowed (from the restrictions of TW2).

It is also interesting to see that since the reduced waiting time cost gives us schedules with more coordinated requests (more shared vehicles), this also gives increased excess ride times.

4.6 Changes to excess ride time costs

As expected, the total excess ride time as well as the worst excess ride time increases with the lower cost, as can be seen in Table 7, but it does not seem to affect the other evaluation criteria. This is probably due to that the high cost for waiting time and the narrow time windows used in the base scenario do not give us the possibility to benefit from lower excess ride time costs.

Analysis of the objective function value reveals an interesting behavior. It might look a bit strange that the total objective function value is slightly higher than in the base scenario, since the only change made is that a cost parameter in the objective function is decreased. However, here we must keep in mind that we work with a dynamic setting where the change can make it more beneficial to insert some requests into the schedule in a specific way, but where it is later no longer possible to insert other requests in as good ways as in the base scenario. By this reason, the

Table 7 Results of excess ride time cost simulations

Evaluation criterion	ETC = 50 % (%)	*p* value	ETC = 150 % (%)	*p* value
No. of vehicles (no. of started shifts)	−1.3	0.35	−2.0	0.16
Total objective function value	0.9	0.03	0.3	0.41
Total vehicle ride time	0.5	0.28	0.5	0.41
Total vehicle ride time with passengers	0.1	0.81	−0.4	0.23
Total vehicle ride time without passengers	1.0	0.18	1.5	0.14
Total request excess ride time	16.0	0.00	−12.8	0.01
Worst excess ride time	26.6	0.02	8.3	0.37
Total deviation from promised pick-up time	−1.6	0.77	−9.1	0.11
Worst deviation from promised pick-up time	4.0	0.19	0.1	0.99

effect on the objective function value cannot be generalized to be valid for all possible instances.

Actually, the same kind of effects can be seen in the results of the tests where TW2 is increased (Sect. 4.3). This tells us that to be able to conclude anything regarding the objective function value of such changes we would need to study several scenarios based on different input, such as several days of requests. And even then it is difficult to draw any conclusions regarding the objective function value from tests where changes are made to waiting time costs and excess ride time costs, since in these tests we change parameters that directly contribute to the objective function value.

5 Discussion and comparison of TW1 results

In this section, our results from the simulations of TW1 are discussed and compared to the results obtained in Quadrifoglio et al. (2008), performed on an American (US) case. They have studied the effects of time window sizes for a demand responsive service operating in Los Angeles.

Our studies have a lot in common. Also they use simulation as their evaluation method. Regarding the input, the Los Angeles service has about twice as many requests per day as our Norrköping case, but both data sets consist of 25 % requests involving wheelchair customers. The paratransit provider of the Los Angeles service uses a time window size of 20 min (and typical in Sweden is 30 min for insertion of the request, but then it is only allowed to postpone this by at most 10 min). As mentioned earlier, about 58 % of the requests in our input where known at least a day in advance. In the Los Angeles case that number is 35 %.

Some important differences exist. Quadrifoglio et al. (2008) generate various scenarios based on 10 days historical data, and consider no-shows and cancellations. Another important issue regards the vehicle fleets. In the Los Angeles case, 12 h shifts are used, and they can be extended to double shifts if needed. In our case, vehicles are required from taxi companies when needed, but maximum for 4 h per shift. When the requests are planned, i.e. inserted into the schedule, one vehicle at a
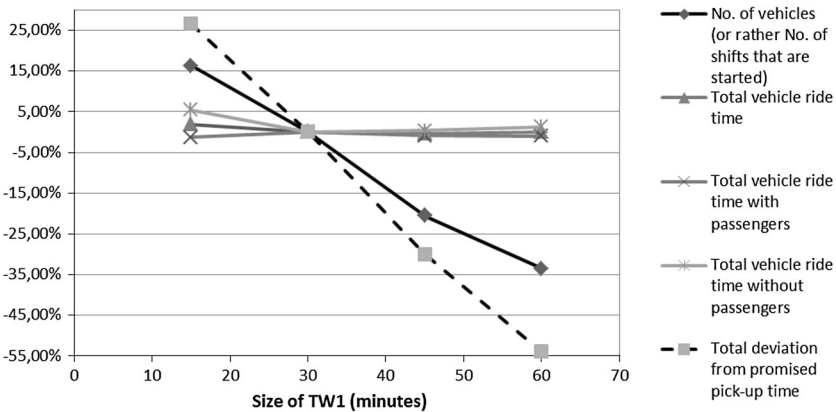
Fig. 7 Description of the effects occurring to some of the evaluation criteria when changing the value of TW1

time is filled up in the Los Angeles case, while in the Norrköping case it is always inserted in the best way according to the objective function described in Sect. 3.1. Another important difference is that in the Norrköping case, we have very high costs for customer waiting times (and also costs for excess ride time), while in the Los Angeles case no such costs are used.

With these differences and similarities in mind, the results of the two studies can now be compared. The results of Quadrifoglio et al. (2008) show that in their study, the total mileage of the vehicles, the dead-heading mileage and the number of vehicles needed can be estimated by decreasing linear functions, while the mileage with passengers onboard increases linearly. These kinds of results are very nice to get since they are easy to describe to practitioners.

The results of these evaluation criteria from the Norrköping case are presented in Fig. 7. As can be seen in the figure, simulations have also been performed for $TW1 = 45$, in addition to the results presented in Sect. 4.2. The difference between the results of using TW1 of 15 min or of 30 min, gives the same kind of relation/ change as in the results from the Los Angeles case. However, the effects are smaller than in the Los Angeles case and when increasing TW1 further (to 45 and 60 min) any additional effects cannot be seen, except for the number of vehicles used which continue to decrease. Instead, the service indicators; waiting time and excess ride time continue to decrease (only waiting time is presented in the figure).

So, the differences between our results are due to how the services are operated, i.e., what costs are used in the objective functions when planning the two different services.

6 Conclusions

In many paratransit systems, no costs regarding customer discomfort (such as waiting time) are considered in the planning of the requests. In such systems a

request can be scheduled anywhere within the specified time window, in the way most suitable for the overall schedule. In the studied case we have included costs regarding excess ride time of each customer and regarding how much the actual pick-up is delayed from the pick-up time agreed upon with the customer (promised pick-up time). The vehicles used in the studied case only inflict costs for the time that a vehicle is used and not as in many other studies where vehicles are often assumed to cost the whole time, no matter if it is used only for serving one request, or used the entire day.

Our results show that when changes are made that increase the possibility to find "better" solutions (e.g. relaxing a set of service constraints), the customer related costs have decreased more than the vehicle related costs. This is of course due to how the relation is between vehicle related costs and service related costs in the base scenario.

The results also show that there is no reason to increase TW2 without decreasing the waiting time cost. Lower waiting time costs in combination with larger values for TW2 allows for better conditions to benefit from the replanning possibility. This would of course give many passengers longer waiting times, which is not really seen as an acceptable solution nowadays. However, many paratransit operators are investigating the possibility to use a "call-back function" meaning that the system automatically calls back to the customer at a specified time before the promised pick-up time, and either confirm the pick-up time or informs that there is a change of plan and gives the new pick-up time. According to our results one must reconsider what customer related costs are used before introducing such a function to the system.

References

Bailey W, Clark T (1987) A simulation analysis of demand and fleet size effects on taxicab service rates. In: WSC '87: proceedings of the 19th conference on Winter simulation, pp 838–844

Cordeau J, Laporte G (2007) The dial-a-ride problem: models and algorithms. Ann Oper Res 153:29–46

Deflorio F (2011) Simulation of requests in demand responsive transport systems. IET Intel Transport Syst 5:159–167

Fu L (2002) A simulation model for evaluating advanced dial-a-ride paratransit systems. Transp Res Part A 36:291–307

Häll C, Peterson A (2013) Improving paratransit scheduling using ruin and recreate methods. Transp Plan Technol 36:377–393

Häll C, Högberg M, Lundgren J (2012) A modeling system for simulation of dial-a-ride services. Public Transport 4:17–37

Parragh S, Doerner K, Hartl R (2008) A survey on pickup and delivery problems: part II: transportation between pickup and delivery locations. J für Betriebswirtschaft 58:81–117

Quadrifoglio L, Dessouky M (2008) Sensitivity analyses over the service area for MAST services. In: Hickman M, Mirchandani P, Voß S (eds) Computer-aided systems in public transport, Lecture Notes in Economics and Mathematical Systems 600. Springer, Heidelberg, pp 419–432

Quadrifoglio L, Dessouky M, Ordóñez F (2008) A simulation study of demand responsive transit system design. Transp Res Part A 42:718–737

Shinoda K, Noda I, Ohta M, Kumada Y, Nakashima H (2004) Is dial-a-ride bus reasonable in large scale towns? Evaluation of usability of dial-a-ride systems by simulation. In: Kurumatani K, Chen S, Ohuchi A (eds) Multi-agent for mass user support, Lecture Notes in Computer Science, vol 3012. Springer, Berlin, pp 105–119

Wilson N, Sussman J, Goodman L, Higonnet B (1969) Simulation of a computer aided routing system (CARS). In: Proceedings of the third conference on applications of simulation, pp 171–183