Policy Recommendations for a Shipment-Consolidation Program

James K. Higginson and James H. Bookbinder

It is well appreciated by logistics professionals that combining several small orders into one larger load can reduce the total transportation cost to the shipper. Logistics specialists all recognize that such a decision to consolidate must take into account the shipper’s increased inventory carrying cost, and the possible impact on service to the customer (consignee), when an order is held for a longer time before dispatch. However, much of the literature on shipment consolidation is descriptive, not sufficiently quantitative to help a company develop a consolidation policy.

This article is addressed to that end. Through a simulation model, we consider three general strategies for dispatch of a consolidated shipment. These strategies are referred to as a Time Policy, a Quantity Policy, and a Time-and-Quantity policy. Simulation results are discussed and understood in terms meaningful to distribution practitioners. For a large range of order-arrival rates and maximum holding times, we compare these policies on the basis of cost per load, cost per hundredweight, and average order delay. Appropriate consolidation policies are then suggested for various situations.

Management must make a number of decisions in developing a shipment-consolidation program. These decisions can be generalized as follows:

- *What* will be consolidated? Which customer orders will be consolidated and which shipped individually?
- *When* will customer orders be released? What event(s) will trigger the dispatch of a consolidated vehicle load?
- *Where* will the consolidation be done? Should consolidation take place at the factory, on a vehicle, at a warehouse or terminal, etc.?
- *Who* will consolidate? Should consolidation be performed by the manufacturer, shipper, customer, carrier, or a third party?
- *How* will consolidation be carried out? Which specific consolidation techniques will be used?

Excellent overviews of the shipment-consolidation problem are given by Tyworth et al.\(^1\) and by Newbourne.\(^2\)

This paper deals with the “When” question: how long should customer orders be held and/or what quantity should be accumulated before a consolidated load is released? Figure 1 presents a diagram of this decision problem for a newly arrived customer order (i.e., a new request for particular merchandise). We defer discussion of previous research on the “When?” question to a later section of the article.
Determining when shipments should be dispatched can be viewed as a two-stage process. First, management must select a shipment-release policy; that is, a general approach for guiding the decision. Second, the chosen shipment-release policy must be operationalized. The result of that stage is a set of shipment-release parameters.

The decision of when to dispatch a consolidated load may be based on a large variety of factors. This paper examines a special class of shipment-release policies—those based only on elapsed time and accumulated quantity. Elsewhere, we examine methods for determining shipment-release parameters for these policies.3,4

SHIPMENT-RELEASE POLICIES

There are three commonly used shipment-release policies. A time policy dispatches each order at a predetermined shipping date, whether or not it is consolidated. This approach sometimes is referred to as “scheduled shipping.” Under a quantity policy, all orders for a particular destination are held and shipped when a minimum consolidated weight is reached. Lastly, a time and quantity policy holds all orders for a particular destination until the earliest of: (i) a predetermined shipping date, or (ii) the accumulation of a minimum weight or volume. If the latter occurs first, the orders are dispatched before the prespecified release date.

We remark that a quantity policy is probably the easiest to use of the three shipment-release policies. A quantity policy is a discrete-time policy; the “state of the system” is checked only when a new order arrives. The time and time-and-quantity policies are analogous to a continuous-time review system.

The focus of a time policy often is customer service, with the shipping date set to meet consignee requirements. Conversely, the target consolidated weight of a quantity policy usually is determined by considerations of cost. Intuitively, a time-and-quantity policy might be expected to exhibit the best characteristics of both a time policy and a quantity policy. This is not necessarily true, as will be seen later.

A time policy, as well as the time component of a time-and-quantity policy, can be approached in two ways. With a maximum holding time (or oldest order) approach, the shipping date is determined by the order that has been waiting longest. Thus, the time component dictates that a consolidated load be dispatched when the oldest order reaches a certain age. The next accumulation cycle begins upon arrival of the first order after this dispatch.
The time component of a scheduled shipping and volume approach may or may not consider the age of the oldest order. For example, the accumulation cycle may begin immediately after the dispatch of a consolidated load or when the first order of a new cycle arrives. If the quantity component is ignored, beginning a new cycle immediately upon load dispatch is similar to releasing shipments at set intervals, with the possibility of cancelling a load if the accumulated weight is too small to be economical. Powell\textsuperscript{5} and Powell and Humble\textsuperscript{6} studied this case through bulk-queuing theory for passenger vehicle dispatching.

Jackson surveyed the use of these three policies by practitioners. A time policy was found to be the most frequently applied (36\% of respondents), but the differences in percentage-usage between the three policies was not large.

Other authors, when investigating the effect of varying consolidation parameters, have commented on the impact of shipment-release policy. Typically, however, these studies have focused on factors other than shipment-release timing, and results are interdependent. For example, although Cooper\textsuperscript{7,8} included both one- and four-day holding times in her simulation, the goal was comparison of direct shipping versus use of warehouses, not the testing of shipment-release policies. Jackson\textsuperscript{9} did compare a time policy with a time-and-quantity policy. Unfortunately, he also varied the number of consolidation points and the system volume; thus his results regarding shipment-release timing are intertwined with those for other parameters.

In this article, we present a simulation model designed expressly to examine the effect on mean per-unit cost and mean order delay of the three shipment-release policies. The simulation results also will be used elsewhere\textsuperscript{10} to test a sequential decision model for timing the dispatch of consolidated loads.

SIMULATION COMPARISON OF SHIPMENT-RELEASE POLICIES

Our simulation was a simple discrete-event model. Customer orders, each weighing a variable amount and arriving at a random time, were generated by computer and accumulated until a target weight or time was reached. All waiting orders then were dispatched, statistics updated, and the accumulation cycle restarted. Because the system emptied with each dispatch, output collection could begin immediately without determination of a steady state. The next sections discuss some modeling considerations.

Simulation Parameters

The simulation assumed common carrier transportation. It is thus important to consider the possibility of phantom freight: the ability to declare heavier weights than actually exist, to push the weight of a load into a heavier weight bracket and qualify for a lower freight rate.\textsuperscript{11,12} The minimum weight (which we will call "WBT") at which this strategy is cost-effective equals the minimum volume weight ("MVT") times the ratio of the volume freight rate $f_v$ and the non-volume freight rate $f_n$; that is, $WBT = MVT \left( \frac{f_v}{f_n} \right)$, where WBT $\leq$ MVT and $f_n = f_n$. At that weight, WBT $f_n = MVT f_v$. The concept of over-declaring the shipment weight (also known as the "bumping clause") is discussed in greater detail in a companion paper by one of us.\textsuperscript{13}

Truck capacity restrictions were ignored, consistent with the popular assumption that additional common-carrier vehicles always are available. The minimum volume weight (MVT) was set at MVT = 20,000 pounds, with volume and non-volume rates of $f_v = \$2.25$ per cwt. and $f_n = \$3.00$ per cwt. respectively. This gives a WBT weight of 15,000 pounds.

Both a quantity policy and a time-and-quantity policy require specification of a target accumulated weight. We arbitrarily selected target weights of 12,500 lbs., 15,000 lbs., 17,500 lbs., 20,000 lbs., and 22,500 lbs., then used the economic shipment weight (ESW) formula to calculate the corresponding order arrival rate. The economic shipment weight\textsuperscript{14} plays the same role here as the EOQ in purchasing management. The ESW is the load size that minimizes the per-order sum of transportation and inventory-holding costs:

$$
ESW = \sqrt{\frac{2 \times F_L \times E[W]}{r_w}}
$$

where $\lambda$ is the order arrival rate, $F_L$ is the sum of all fixed costs associated with a vehicle load, $E[W]$ is the expected weight per customer order, and $r_w$ is the variable cost of carrying inventory per unit weight per time period.

Inventory-holding cost was set at $r_w = \$0.25$ per cwt. per day, and a fixed cost per load of $F_L = \$30$ per dispatch was levied. These cost parameters were constant throughout the simulation. Setting the above ESW expression equal to the selected target weights determined the scenarios we investigated. Table 1 shows these cases in terms of the order arrival rates, and the relationships between the
target weights and WBT and MWT. Note that these figures correspond to an average order size of \( E[W] = 2,000 \) lbs.

**TABLE 1**

<table>
<thead>
<tr>
<th>Target Weight</th>
<th>Comment</th>
<th>Order Arrival Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>12,500 lbs.</td>
<td>less than WBT</td>
<td>3.26 orders per day</td>
</tr>
<tr>
<td>15,000 lbs.</td>
<td>equal to WBT</td>
<td>4.69 orders per day</td>
</tr>
<tr>
<td>17,500 lbs.</td>
<td>between WBT and MWT</td>
<td>6.38 orders per day</td>
</tr>
<tr>
<td>20,000 lbs.</td>
<td>equal to MWT</td>
<td>8.33 orders per day</td>
</tr>
<tr>
<td>22,500 lbs.</td>
<td>greater than MWT</td>
<td>10.55 orders per day</td>
</tr>
</tbody>
</table>

MWT is the minimum weight needed to qualify for a volume freight rate; WBT (<MWT) is the weight at which the “bumping clause” comes into effect.

For the time policy and the time-and-quantity policy, an oldest order approach was used to set the maximum holding time: consolidated loads were dispatched when the order waiting the longest had reached a specified maximum delay. These times were arbitrarily selected after considering the expected time to accumulate the target weight given the arrival rate. We tested maximum holding times of 0.75 days, 1.0 days, 1.5 days, and 2.0 days.

**Modeling Customer Order Arrivals and Weights**

Our simulation assumed that order arrivals follow a Poisson process with an arrival rate of \( \lambda \) orders per day (hence the interarrival times are exponentially distributed). Masters\(^{15}\) modeled interarrival times between orders as a uniform distribution, while Cooper\(^{16}\) used an exponential distribution, and Jackson\(^{17}\) and Ha, Khasnabis, and Jackson\(^{18}\) used empirical distributions.

Very little guidance exists in the academic literature as to appropriate frequency distributions of customer-order weights. Such a distribution will vary between and among products, shippers, carriers, and purchasers.

Masters modeled customer order weight as a normal distribution with coefficient of variation (CV) of 1/2 (CV equals the standard deviation divided by the mean). Cooper and Ha, Khasnabis, and Jackson used truncated normal distributions with CV = 1. Jackson and Closs and Cook\(^{19}\) employed empirical data for customer-order weight, but gave no further details.

Fitting of theoretical probability distributions to empirical order weights was attempted by Akaah and Jackson.\(^{20}\) Unfortunately, their analysis was limited to the normal, uniform, and Poisson distributions. Although only ten of their forty sets of weights fit one of these distributions, no better-fitting distributions were suggested or tested.

Jackson’s\(^{21}\) simulation used empirical order weights from a medium-size national package goods distributor. Consistent with that data, we utilized an unshifted gamma distribution (denoted as \( \text{Ga}(\alpha,\beta) \)) for modeling order weight (see also Higgins\(^{22}\)). Our reasoning included visual comparison with Jackson’s empirical plot as well as analysis of other data sets from industry. Many of these empirical data exhibited skewness towards lower weights. Because this is commonly observed in reality, we employed the value \( \alpha = 2 \) to obtain a distribution that is more concentrated toward lower weights. Indeed, shipment-consolidation is important in logistics practice precisely because the larger weights are not prevalent.

For simplicity, we selected a mean order weight of 2,000 pounds. The expected value of the gamma distribution is \( E[W] = \alpha \beta \), hence \( \beta = 1,000 \). The standard deviation is then \( \sigma = \beta \sqrt{\alpha} = 1,414 \) pounds. The mode (the weight that occurs most often) is \( \beta(\alpha-1) = 1,000 \) pounds, only slightly higher than Jackson’s mode.

**SIMULATION RESULTS**

Output data was analyzed using the method of batch means.\(^{23}\) Batches of customer orders were generated prior to the simulation, then independently processed according to a time policy, a quantity policy, and a time-and-quantity policy. Undispatched orders remaining at the end of the simulation were deleted so that only complete consolidation cycles were considered.
Figures 2 and 3 illustrate results relating to per-cwt. cost, while Figures 4 and 5 give results for mean order delay. Because a quantity policy is not affected by maximum holding time, per-cwt. cost and mean order delay are identical on these graphs for this strategy. Figures 6 and 7 restate our simulation results in terms of order arrival rate, rather than holding time as used in Figures 2 through 5.

Significance testing of the differences between policies was done using a paired-t confidence interval at the 90% level of confidence. Unless stated otherwise, the differences between policies were found to be statistically significant. The main exception was the comparison of a quantity policy and time-and-quantity policy. For arrival rates of \( \alpha \geq 8.33 \), differences were not significant for per-unit cost with maximum holding time of 1.5 days, and for both per-unit cost and mean delay with holding time of 2.0 days. Under these conditions, the time restriction of a time-and-quantity policy is more active than the quantity component in determining the timing of load dispatches. As well, some differences were not significant when the plot of results for one policy crossed that of another (for example, per-unit cost of a time policy versus time-and-quantity policy with arrival rate \( \alpha = 8.33 \) and holding time of 1.5 days).

Figures 2 and 3 do not always show a quantity policy as the lowest per-unit cost strategy. This is because the discounts for larger load weights were not considered when selecting the target weights of this policy. Although the economic shipment weight formula used to derive the target weights yields the minimum cost weight, when weight breaks are considered, a lower cost weight may exist. For example, with freight rate weight breaks, the deterministic per-unit cost with arrival rate \( \alpha = 6.38 \) would be about $2.58 per cwt., as compared to the mean cost of $2.74 actually reported in Figures 2 and 3.

Figures 2 through 5 illustrate the important interaction between order arrival rate and maximum holding time. Low order arrival rates coupled with short holding times may yield different preferred policies than will high order arrival rates and long holding times. Moreover, no one policy yields the lowest per-cwt. cost for all order arrival rates. Thus, there is no one best order-release policy, although a time-and-weight policy consistently yielded the smallest mean delay per order. The choice and performance of a shipment-release strategy is heavily dependent on the order arrival rate and the length of time customers are willing to wait for shipment. This is especially important when the order arrival rate is low.
FIGURE 3
COMPARISON OF SHIPMENT-RELEASE POLICIES:
MEAN COST PER CWT
HOLDING TIME = 2.0 DAYS

FIGURE 4
COMPARISON OF SHIPMENT-RELEASE POLICIES:
MEAN ORDER DELAY
HOLDING TIME = 1.0 DAYS
FIGURE 5
COMPARISON OF SHIPMENT-RELEASE POLICIES:
MEAN ORDER DELAY
HOLDING TIME = 2.0 DAYS

FIGURE 6
COMPARISON OF SHIPMENT-RELEASE POLICIES:
MEAN COST PER CWT
ARRIVAL RATE = 10.55 ORDERS PER DAY
The following paragraphs discuss these results with regard to specific policies. Following this, we make more detailed comments as to the preferred policy under various combinations of order arrival rate and maximum holding time.

Quantity Policy

The cost-performance of a quantity policy, relative to one that considers holding time, depends on the selected order holding time. A quantity policy will be more expensive than a time policy if the latter has a holding time long enough to accumulate loads sufficiently large to benefit from volume transportation discounts. However, a holding time such as 0.75 days is so short that time-based strategies cannot do this, and are outperformed cost-wise by a quantity policy. As the maximum holding time increases, a time policy produces cost results comparable to a quantity policy because the holding time now is sufficiently long that a time policy can accumulate load sizes comparable to that of a quantity policy. A quantity policy also had the smallest coefficient of variation of per-load cost in our simulation.

Time Policy

Our simulation shows the danger of a pure time policy. A short holding time coupled with a low order arrival rate will result in frequent small loads. Long holding times will produce excessive load sizes, with benefits from transportation savings overwhelmed by inventory costs. In both cases, the per-unit cost will be larger than that of other policies. Moreover, because the maximum holding time has a direct effect on mean order delay, a poor choice can yield excessive delays, as seen in Figure 5 for high arrival rates with a holding time of 2.0 days.

A time policy consistently yielded the greatest coefficient of variation of per-load cost, sometimes twice that of a time-and-quantity policy, and as much as five times that of a quantity policy. As well, a time policy always produced loads at least as large as (and frequently larger than) those of a time-and-quantity policy. The difference in load size between the two policies was not great (though statistically significant) for low arrival rates and short holding times, but was dramatic for high arrival rates and long holding times, where a time policy produced loads as much as 44% bigger. Although the difference in load sizes also depends on the order arrival rate, the maximum holding time has a much greater impact.

With low arrival rates and short holding times, the larger loads of a time policy resulted in lower per-unit cost than did a time-and-quantity policy. However, with
high arrival rates and long holding times, the cost performance of a time policy suffered because transportation savings from volume loads were overwhelmed by inventory-holding costs.

**Time-and-Quantity Policy**

It might be expected that a time-and-quantity policy would exhibit the best features of both a quantity policy and a time policy. Our results show that this may not always be true, depending on the basis of comparison. When the holding time is short, the target weight is not operative because there is insufficient time to reach that weight. A time-and-quantity policy then performs much like a time policy, resulting in higher per-unit cost than does a quantity policy. When the maximum holding time is long, the target weight is reached before the time limit is reached, and the strategy acts like a quantity policy. We realized this by comparing per-unit cost for a quantity policy and a time-and-quantity policy for arrival rates of 8.333 and holding times of 2.0 days (Figure 3) and 1.5 days; the differences between the two polices for these \( T_{\text{max}} \) values are not statistically significant. This can also be seen by comparing the relevant portions of Figure 6.

Intuitively, a time-and-quantity policy should never be more expensive than a time policy, nor should it be cheaper than a quantity policy. Our results agree with this thinking on a cost per-load basis; this is not true on a per-cwt. basis. For example, for low order arrival rates, a time policy yields greater per-load cost than does a time-and-quantity policy. However, the load produced by a time-and-quantity policy is never larger than that of a time policy because the former policy is constrained by a target weight; a time policy will allow shipments greater than the target weight. Thus, with low arrival rates and short holding times, shipments under a time-and-quantity policy often will not be sufficiently large to qualify for volume freight rates, while the greater loads accumulated by a time policy may. As a result, on a cost per-cwt. basis, a time-and-quantity policy may be more expensive than a time policy.

In terms of mean order delay, however, a time-and-quantity policy outperforms the other two strategies, chiefly for the same reasons that it performs poorly with regard to per-unit cost. When the order arrival rate is low, the time portion of the strategy is active, and loads are dispatched without waiting for the target quantity to be attained. When the holding time is long, the target quantity is attained first, thus avoiding the excessive waiting times of a time policy.

**Comparison of Our Simulation Results With Those in the Literature**

The impact on cost and delivery time from changes in \( T_{\text{max}} \) were noted by Masters for a time policy and by Cooper for a time-and-weight policy. Both concluded that longer holding times reduce transportation costs and increase mean delivery time, although the latter increase is a fraction of the increase in \( T_{\text{max}} \). These conclusions agree with ours.

Jackson examined the impact on shipment cost and order cycle time from changes in the maximum holding time and system volume. He also compared the cost and time performance of the scheduled shipping approach (time policy) and the scheduled-shipping-and-volume approach (time-and-quantity policy). His simulation model differs from ours in several respects. Jackson varied such factors as the number of consolidation points and the system volume; direct shipping without consolidation also was tested. Transportation time, based on empirical results of Piercy, was included in his calculation of order delay. Like Masters and Cooper, however, inventory-holding costs were not considered.

Jackson found that low volume systems suffer from greater transportation costs and longer, more variable order delays. This agrees with our results when the order arrival rate is low. As well, we agree with his conclusion that longer holding times reduce transportation cost, but add that if the dispatch of a shipment is determined solely by target weight or quantity, longer holding times have no effect on transportation cost.

Jackson noted that the combination of long holding times and large system volumes (high order arrival rates) results in lower costs. This is true if inventory-holding costs are ignored, as Jackson did. If they are considered, long holding times may cause transportation savings to be overwhelmed by inventory-holding costs, as seen in Figure 6 for a time policy.

We also add that, for any strategy that includes a target consolidated weight, if either the arrival rate or the holding time is sufficiently large that most shipments move under the lowest volume rate, further increases in either parameter will not change per-cwt. cost appreciably. This can be seen by comparing per-cwt. cost for holding times of 1.0 days (Figure 2) and 2 days (Figure 3) with order arrival rates of 8.333 and 10.546.
Jackson also noted that a time strategy was cheaper and slower than a
time-and-weight strategy for high volume systems, but comparable for small volumes.
Our results agree. However, for low arrival rates, the difference in mean order
delay between the two policies becomes considerably greater as the maximum holding
time increases.

Powell\textsuperscript{28} compared vehicle dispatch strategies through analysis of batch
arrivals/bulk service queues. He included a service measure calculated as:

$$W_{95} = W_q + 1.645 \text{var}(W_q)^{1/2}$$

where $W_{95}$ is the 95th percentile of waiting time, $W_q$ is the mean waiting time,
and the constant 1.645 is derived by assuming that the distribution of waiting times
is approximately normal. His assumptions differ from ours (for example, order arrivals
occurred in batches, and scheduled vehicles could be cancelled if the load was
insufficient), so direct comparison of results is not straightforward. However, he
found that shorter maximum holding times were more expensive than longer ones,
and that a weight strategy dominated a time strategy according to his $W_{95}$ service
measure. This latter conclusion does agree with our results for long holding times,
or for short holding times with very large arrival rates.

On the whole, our findings support other research in illustrating how complex
the choice of a shipment-release policy can be. We next offer some specific comments
for deciding on such a strategy, while the final section makes some general
conclusions.

### WHICH POLICY IS BEST?

Given certain combinations of order arrival rate, minimum-cost quantity, and
maximum holding time, which policy is preferred? Figures 6 and 7 restate our
simulation results for a particular order arrival rate, while Table 2 summarizes these
figures and others in terms of dominated policies. As this table shows, for many
combinations of arrival rate and holding time, the best policy will depend on
management's objectives with regard to cost and customer service.

More exact decision rules suggesting the best policy would be helpful.
Developing such rules is difficult: parameters may take a wide range of (not

<table>
<thead>
<tr>
<th>Holding Time</th>
<th>Order Arrival Rate (Orders Per Day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda=3.26$</td>
<td>$\lambda=6.38$</td>
</tr>
<tr>
<td>0.75 days</td>
<td>No clear choice</td>
</tr>
<tr>
<td>(0.245)</td>
<td>(0.479)</td>
</tr>
<tr>
<td>1.0 days</td>
<td>Time policy dominates</td>
</tr>
<tr>
<td>(0.326)</td>
<td>(0.638)</td>
</tr>
<tr>
<td>1.5 days</td>
<td>Quantity policy dominates time policy</td>
</tr>
<tr>
<td>(0.489)</td>
<td>(0.957)</td>
</tr>
<tr>
<td>2.0 days</td>
<td>Quantity policy dominates time policy if target weight equals minimum volume weight</td>
</tr>
<tr>
<td>(0.652)</td>
<td>(1.876)</td>
</tr>
</tbody>
</table>

### Definitions:

For a given combination of arrival rate $\lambda$ and holding time, a particular
policy (say $P_1$) dominates another policy ($P_2$) when both the cost and
delay performances of policy $P_1$ are at least as good as those for
policy $P_2$, and one of these performances is better for $P_1$ than for $P_2$.

"No clear choice" means that the choice of shipment-release policy
under this combination of arrival rate and holding time will depend on
management objectives of cost and customer service.
necessarily optimal) values. For example, management may prefer a smaller-than-optimal target weight, trading higher per-unit cost for improved customer service.

We tested the following quantitative decision approach. First, given the cost parameters in the simulation, we determined the minimum-cost consolidated weight for each arrival rate \( \bar{a} \). We next calculated \( E[\%] \), the percentage of this minimum-cost weight that would be expected to be consolidated within each holding time \( T_{\text{max}} \); that is:

\[
E[\%] = E[\text{percent of minimum-cost weight accumulated in holding time}]
\]

\[
= T_{\text{max}} \times E[W] / \text{minimum-cost weight}
\]

where \( E[W] \) is the expected weight of an order.

Values of \( E[\%] \) for the twelve arrival-rate/holding-time combinations simulated are given in parentheses in Table 2. Comparing these values to the summary of results given in that table yields the observations in Table 3.

Obviously, clearer definition of boundaries is possible. We also stress that because \( E[\%] \) was calculated using the minimum-cost shipment weight, these conclusions are only valid if the selected target weight equals this minimum-cost weight.

The following example illustrates the decision approach of Table 3.

Suppose that the minimum-cost shipment weight is found to be 20,000 pounds (see Higginson29 for a discussion of this calculation). Management estimates that the maximum waiting time acceptable to its customers is \( T_{\text{max}} = 3 \) days, the arrival rate is \( \bar{a} = 4 \) orders per day, and the mean order weight is \( E[W] = 3,000 \) pounds.

The resulting value of \( E[\%] \) is 1.8. From the above observations, we conclude that either a quantity policy or a time-and-quantity policy should be adopted (a quantity policy may be preferred simply because it is easier to use). Indeed, if maximum holding time were reduced to \( T_{\text{max}} = 2.345 \) days, \( E[\%] \) would equal 1.407, and, according to Table 3, the choice of shipment-release policy would not change.

### TABLE 3

**DECISION RULES AND PREFERRED POLICY IN TERMS OF CALCULATED \( E[\%] \), THE FRACTION OF THE MINIMUM-COST WEIGHT THAT, ON AVERAGE, WOULD BE AVAILABLE FOR CONSOLIDATION WITHIN A HOLDING TIME \( T_{\text{max}} \)**

<table>
<thead>
<tr>
<th>( E[%] ) Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 \leq E[%] \leq 0.703 )</td>
<td>No clear choice, preferred policy dominates other policies regarding both cost and mean order delay. Preferred policy will depend on management objectives regarding cost and customer service.</td>
</tr>
<tr>
<td>( 0.938 \leq E[%] \leq 1.276 )</td>
<td>Time policy is not recommended because it is dominated by both a quantity policy and a time-and-quantity policy. Differences between a quantity and a time-and-quantity are sufficiently large that the preferred policy will depend on management objectives regarding cost and customer service.</td>
</tr>
<tr>
<td>( E[%] \geq 1.407 )</td>
<td>Either a quantity or time-and-quantity policy is recommended: performance is similar or statistically the same for both.</td>
</tr>
</tbody>
</table>

If the volume of orders (as reflected in \( \bar{a} \)) decreased so that the \( E[\%] \) value fell below 0.703, management then should reconsider its choice of shipment-release policy in light of its goals regarding the trade-off between cost and customer service.

### CONCLUSIONS

Two questions can be asked of our simulation. First, how would the results differ if we had simulated private carriage instead of common carriage? For a private carrier, cost savings from consolidation occur from spreading of fixed costs, mainly those related to transportation, over larger loads. However, as the load size grows, the marginal change in per-unit transportation cost decreases. Thus, policies that hold orders for long periods so as to accumulate the common carrier's minimum
volume may not be as effective under private carriage. Of course, results will depend on cost parameters and the minimum-cost shipment quantity.

Second, we noted that carrier volume discounts were not considered when setting the target weights of a quantity and a time-and-quantity policy. How would our results be affected if these discounts were included when determining target quantities? Analysis elsewhere\(^\text{30}\) shows that the target weight should equal the minimum volume weight (20,000 pounds in our simulation) for all arrival rates less than 8.33 (recall that the target weight was 20,000 pounds for \(\bar{d} = 8.33\)). Thus, results for \(\bar{d} = 8.33\) and \(\bar{d} = 10.55\) would not change. For each arrival rate, a quantity policy would result in the lowest per-cwt. cost, as seen in all figures for \(\bar{d} \geq 8.33\). With a quantity policy, mean order delay would increase for all \(d < 8.33\) because the target weight would be larger. The amount of change in delay performance would be greatest for low arrival rates, but would not change the relative rankings of the three policies: a quantity policy yields the longest mean order delay even with the smaller target quantities simulated. The rankings might change for arrival rate \(\bar{d} = 6.38\) because a quantity policy produces a mean delay more comparable to the other policies.

From our simulation results, we can make the following conclusions regarding the relative cost and delay performance of the three shipment-release policies.

**Quantity Policy**

Cost: A quantity policy will yield the lowest per-unit distribution cost, assuming that the lowest-cost quantity is selected. If the lowest per-unit cost quantity is not selected, the performance of this policy relative to that of the time policy will depend on the holding time and the order arrival rate. For example, a weight strategy will be bettered by a time strategy if the latter has a holding time long enough to accumulate large loads.

Delay: A quantity policy may or may not outperform a time policy, depending on the holding time selected, but will never perform better than a time-and-quantity policy.

**Time Policy**

Cost: A time policy can be dangerous cost-wise. A short holding time and small order arrival rate will result in frequent small loads and increased cost. Long holding times will produce excessive loads, with benefits from transportation savings overwhelmed by inventory costs. As well, this policy yields the largest variation in per-order cost.

Delay: A time policy may or may not outperform a quantity policy, depending on the holding time selected, but will never perform better than a time-and-quantity policy.

**Time-and-Quantity Policy**

Cost: On a per-unit basis, a time-and-quantity policy will never be cheaper than a quantity policy, and may be more expensive than a time policy because a time-and-quantity policy produces load sizes no larger, and frequently smaller, than those of a time policy.

Delay: A time-and-quantity policy will perform as well or better than both a time policy and a quantity policy for the same reasons that it performs poorly with regard to cost.

From these conclusions, we emphasize that knowledge of the level of service required by customers is crucial in selecting a shipment-release policy. Customer service and order arrival rate must be examined simultaneously, because the value of one can eliminate or reduce significantly the importance of the other in the decision.

After a shipment-release policy has been selected, values for policy parameters must be determined. Approaches to this will be treated in subsequent publications.\(^\text{31}\)

---

Authors' note: The authors' work was respectively supported by the Social Sciences and Humanities Research Council of Canada, through a graduate scholarship, and by the Natural Sciences and Engineering Research Council of Canada, via grant no. OGP 05292.

**NOTES**


3 J. K. Higginson and J. H. Bookbinder, "Probabilistic Models for Order-Holding Times in Freight Consolidation." Working paper, Department of Management Sciences, University of Waterloo. (Submitted for publication, 1993.)


10 Same reference as Note 4.

11 Same reference as Note 1.


13 J. K. Higginson, "Optimal Dispatch Quantities in Goods Transportation." Working paper, University of Waterloo. (Submitted for publication 1992.)

14 Same reference as Note 13.


16 Same references as Notes 7 and 8.
ABOUT THE AUTHORS

James K. Higginson received his bachelor's degree in commerce from McMaster University in Hamilton, Ontario, and his master's and Ph.D. degrees in management sciences/logistics from the University of Waterloo. He taught courses in transportation management and quantitative methods at the University of British Columbia. He is a research associate and consultant to small businesses in Kitchener-Waterloo, Ontario.

James H. Bookbinder is professor of management science at the University of Waterloo. For 20 years he has worked with manufacturing firms, distributors, and transportation carriers on the modeling and analysis of logistics strategies and operations. He is a past-president of the Canadian Operational Research Society and past-chair of the transportation science section of the Operations Research Society of America. He holds an M.B.A. from the University of Toronto and a Ph.D. from the University of California, San Diego. His current research concerns shipment consolidation, distribution requirement planning, intermodality in global supply chains, and transportation aspects of JIT system. He is an associate editor of Operations Research (OR Practice), Naval Research Logistics and Logistics & Transportation Review. He is director of the University of Waterloo's research group in logistics and manufacturing.