Faculty & Research
Working Paper

An Operational Mechanism Design for
Fleet Management Coordination
in Humanitarian Operations

Alfonso J. PEDRAZA MARTINEZ
Sameer HASIJA
Luk N. VAN WASSENHOVE
2010/87/TOM/ISIC

An Operational Mechanism Design for Fleet Management Coordination in Humanitarian Operations

Alfonso J. Pedraza Martinez*

Sameer Hasija**

Luk N. Van Wassenhove***

*   PhD Candidate in Technology and Operations Management at Boulevard de Constance, 77300 Fontainebleau, France. Ph: +33 (0)1 60 71 25 31 Email: alfonso.pedrazamartinez@insead.edu

**   Assistant Professor of Technology and Operations Management at INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. Ph: +65 6799 5388 Email: sameer.hasija@insead.edu

***   Professor of Operations Management, The Henry Ford Chaired Professor of Manufacturing, INSEAD Social Innovation Centre at INSEAD Boulevard de Constance, 77300 Fontainebleau, France. Ph: +33 (0)1 60 72 42 66 Email: luk.van-wassenhove@insead.edu

An Operational Mechanism Design for Fleet Management Coordination in Humanitarian Operations

Alfonso J. Pedraza Martinez

INSEAD, 77305 Fontainebleau Cedex, France {alfonso.pedrazamartinez@insead.edu}

Sameer Hasija

INSEAD, 1 Ayer Rajah Avenue, Singapore, 138676 {sameer.hasija@insead.edu}

Luk N. Van Wassenhove

INSEAD, 77305 Fontainebleau Cedex, France {luk.van-wassenhove@insead.edu}

We study incentive alignment to coordinate operations in humanitarian settings. Our research focuses on transportation, the second largest overhead cost to humanitarian organizations after personnel. Motivated by field research, we study the fleet size problem from a managerial perspective. In terms of transportation, the objective of humanitarian Programs is to have a vehicle available whenever it's needed; the bigger the fleet, the higher the availability (the lower the cost of delay). On the other hand, the bigger the fleet, the higher the fleet cost. Fleet cost is the responsibility of the National Logistics. The different focus of the Programs and the National Logistics creates misaligned incentives that may lead to sub-optimal performance of a decentralized system. At the top of the system, the Headquarter must design incentive mechanisms to balance the operating cost of the fleet with the equity cost represented by cost of delay. The incentive alignment issue is complex in a humanitarian setting as traditional instruments based on financial rewards and penalties are not considered as viable options. The problem is complicated further by information asymmetry in the system due to the disperse geographical location of Programs, National Logistics and Headquarter. We propose a novel mechanism design for the incentive alignment problem where the Programs have private information regarding their true transportation needs. This study contributes to the humanitarian logistics literature and to the incentives in operations management literature.

Key words: Incentives, Humanitarian Logistics, Fleet Management

1. Introduction

The need for humanitarian action has increased dramatically in the last few decades and it is expected to rise significantly in the years to come (Thomas and Kopczak 2005). Organizations

carrying out humanitarian action face serious challenges to deliver the right goods and services to the right people at the right time and at the right cost (Van Wassenhove 2006). Transportation, the second largest overhead cost to international humanitarian organizations after personnel (Disparte 2007), offers interesting research opportunities to improve operational efficiency. Improvements in transportation and fleet management translate in better attention to beneficiaries.

We study the field vehicle fleet management system in International Humanitarian Organizations (IHO). This research focuses on 4x4 light vehicles, the most widely used vehicles to coordinate and execute last mile distribution. The last stage of the humanitarian supply chain, last mile distribution relates to the delivery of humanitarian goods and services to beneficiaries in field operations (Balcik et al 2008).

Our research is motivated by a larger field project to understand field vehicle fleet management in IHO. In the following we will briefly describe the project and use our findings from the field research to motivate the theoretical model that is studied in this paper. The field project includes four large IHO: the International Committee of the Red Cross (ICRC), the International Federation of Red Cross and Red Crescent Societies (IFRC), the World Food Programme (WFP) and World Vision International (WVI). Staff interviews were conducted at the Headquarter in Europe, at national and field office levels in the Middle East and Africa for the various IHO (Pedraza Martinez et al 2010). Typically, the fleet management system in large IHO has three decision making parties during the planning stage of development programs: the Humanitarian Programs, the National Logistics and the Headquarter.

Humanitarian Programs are the primary delivery channel of humanitarian aid for the IHO. Often located in remote areas of developing countries (the field), Programs are service oriented. They provide assistance and help alleviating the suffering of people in the aftermath of disasters (relief). Also, Programs implement activities to improve the quality of life of poor communities (development). Transportation requirements for relief and development are different. Relief Programs assign vehicles according to emergency priorities for disaster assessment, or to coordinate search, rescue and emergency aid distribution operations. Development Programs use vehicles for regular

visits to villages or refugee camps for health care or to coordinate aid distribution. Urgency in development Programs is lower and vehicles are typically assigned to visits in order of requisition. We focus on fleet management in development Programs with big fleets of twenty or more vehicles in the same geographic location. Development Programs are henceforth referred to as Programs.

Some examples of Programs include health, nutrition, water and sanitation. In terms of transportation, the objective of Programs is to have a vehicle available when it is required by the staff to visit beneficiaries. Although speed in demand fulfillment is not necessarily critical, Programs must meet demand in a reasonable time. Due to the long term nature of Programs, visits that cannot be performed on time are accumulated. Programs incur two main costs, the cost of delay and the cost of managing their fleets in the field. Often Program managers are more sensitive to the cost of delay than to the fleet management cost. The bigger the fleet the lower the cost of delay. During the planning stage, Programs state their transportation needs to the National Logistics.

The National Logistics, henceforth referred to as Logistics, compile the Programs' stated transportation needs and make fleet requisitions to the Headquarter. The Logistics are instrumental to the implementation of fleet management policies and procedures. Located in urban centers in the countries of operation, the Logistics are accountable for the running cost of the fleet. The running cost includes maintenance, repairs and fuel consumption. To keep the cost low, during the planning stage the Logistics decide whether to monitor the Programs' stated needs or to approve the stated needs. Monitoring is expensive. It consists of setting up information systems and sending staff to the field to verify the Programs' stated transportation needs. From our field research we learned that Programs keep good data records, usually in printed files or in non-standardized spreadsheets, mainly for audit purposes (Pedraza Martinez et al 2010). Via *in situ* monitoring the Logistics can get an accurate estimation of transportation needs of the Programs before sending fleet requisitions to the Headquarter.

The Headquarter has the function of procuring the fleet requested by the Logistics. Located in the US or Europe, the Headquarter's objective is balancing the cost of delay of last mile distribution and the operating cost of the fleet, i.e. the fleet management cost plus the running cost. During the

planning stage, the Headquarter decides the optimal fleet size to minimize the system's cost. The fleet size includes a fleet buffer to guarantee that a reasonable proportion of visits will not suffer any delay. The fleet buffer determines the service level of the fleet. The bigger the fleet buffer the lower the cost of delay but the higher the operating cost of the fleet.

As described earlier, the organizational structure of the IHO creates a decentralized system with misaligned incentives. Due to the service orientation of the Programs, they may prefer having a larger fleet size than the system's optimal level. The problem of misaligned incentives gives rise to an adverse selection due to asymmetry in information in such a system. While the Headquarter and the Logistics may have probabilistic estimates of the true transportation needs of the Programs, they may not be able to accurately determine the true needs unless they incur an expensive monitoring effort. The monitoring decision is made by the Logistics, that has a sole purpose of minimizing the running cost of the fleet and the cost of monitoring. The Headquarter due to its location cannot monitor the fleet requisition made by the Programs and must trust the Logistics for the stated demand. The Headquarter determines the fleet size based on the transportation needs compiled by the Logistics and a predetermined buffer level to absorb demand and service-time variability in the system. The incentives problem is illustrated by quotes from our interviews. One of the Headquarter staff said:

"*I feel some of our programs have more vehicles than required*"

In fact, one of the Global Fleet Managers we interviewed estimates that their Programs have between 10% and 15% more vehicles than required. A Logistics staff member in Mozambique stated:

"*We do not have effective monitoring tools to control Program fleets*"

In contrast, when we asked about the fulfilment of his transportation needs, a Program relief worker in Mozambique said:

"*We do not have enough vehicles. Often we have to wait too long to have a vehicle available to go to the field*"

We show why, from their own perspective, the three concerns are justified. Then, we focus our effort on the coordination of the fleet management system. But coordinating this system is challenging since standard financial incentives based mechanisms are not considered viable in the humanitarian setting. This is primarily due to the organizational culture of the IHO, wherein the employees are driven by their motivation to serve and not by objectives such as profit maximization (Lindenberg 2001, Manell 2010). E.g., it is almost inconceivable for the IHO to incentivize volunteer medical doctors working in the field by using financial penalties and rewards. We believe that this unique characteristic of humanitarian organizations makes the incentive misalignment issue an extremely interesting research topic, and not just a mere application of the principal agent framework from the economics literature.

Instead of using mechanisms based on traditional financial incentives we propose an operational mechanism to coordinate the system. We use the fleet buffer decided by the Headquarter as an operational lever to achieve truth revelation from Programs in terms of transportation needs. To the best of our knowledge, this is the first attempt of using an operational capacity based mechanism to coordinate a principal-agent system. Our work should specially appeal to the Operations Management community as it showcases the strategic importance of operational design, beyond the objective to achieve tactical efficiency in the humanitarian sector.

The paper is organized as follows. Section 2 positions the paper in the literature. Section 3 describes the IHO fleet management system. Section 4 introduces the operational capacity based mechanism. Section 5 extends the analysis via numerical experiments. Finally, section 6 presents the conclusions and further research agenda.

2. Literature Review

Our paper contributes to the humanitarian logistics literature and to the incentives alignment literature in operations management. The extant literature on humanitarian logistics follows a classical optimization approach. Most of the research examines relief systems for disaster preparedness or for disaster response. Typically, those papers apply operations research techniques to relief operations taking a central planner approach. The objective can be equity or cost-efficiency oriented.

Equity-based objective functions have been studied in terms of time of response and demand fulfilment. Research to minimize the time of response can be found in Chiu and Zheng (2007) and Campbell et al (2008). Research exploring demand coverage include Batta and Mannur (1990), Ozdamar et al (2004), Jia et al (2007), De Angelis et al (2007), Yi and Ozdamar (2007), Saadatseresht (2009) and Salmeron and Apte (2010).

Cost-based objective functions are often represented either via monetary cost or via travel distance. Cost minimization can be found in the work of Barbarosoglu et al (2002), Barbarsoglu and Arda (2004), Beamon and Kotleba (2006) and Sheu (2007). Distance traveled minimization has been explored by Cova and Johnson (2003), Chang et al (2007), and Stepanov and Smith (2009). In their work, Stepanov and Smith also examine an equity based function of time of response. Regnier (2008) models the trade-off between cost and equity in hurricane evacuation operations, also from a central planner perspective.

In contrast to the extant literature in humanitarian logistics we analyze the trade-off equity/efficiency in a decentralized system using a principal-agent approach. In our model the Program has private information regarding its transportation needs. Also, different agents control the two components of the cost function. While the Logistics are responsible for the running cost of the fleet, the Program is responsible for the fleet management cost in the field and incurs the disutility due to delay in reaching their beneficiaries. At the top of the system, the Headquarter's problem is to balance the operating cost of the fleet with the equity cost represented by cost of delay while keeping the system's incentives aligned.

The incentives literature has focused on exploring manufacturing and service operations management in "for profit" settings. Typically, decisions in manufacturing supply chains relate to order-quantity of goods while decisions in service supply chains relate to the capacity of the service system. Most of the mechanisms for supply chain coordination in manufacturing and in service operations management are based on financial incentives.

In this paragraph we briefly summarize some commonly studied financial contracts in manufacturing and service supply chains. This is not a comprehensive list of the vast literature on supply

chain contracts, but it provides the readers a primer on the nature of contracts that have been studied in such settings. In revenue sharing contracts a retailer pays a supplier a wholesale price for each unit purchased plus a percentage of the revenue generated by the retailer (Cachon and Lariviere 2005). Buy-back contracts have a wholesale price and a buy-back price for unsold goods (Pasternack 1985). In Sales-rebates contracts the supplier charges the retailer a per-unit wholesale price but gives the retailer a rebate per unit of goods sold above a predefined threshold (Krishnan et al 2001, Taylor 2002). In quantity discount contracts the retailer receives a discount either on all units if the purchased quantity exceeds a threshold (all-unit quantity discount) or on every additional unit above a threshold (incremental quantity discount) (Corbett and de Groote 2000, Cachon and Terwiesch 2009). In price-discount contracts wholesale prices are discounted on the basis of annual sales (Bernstein and Federgruen 2003). In service systems decisions are based on capacity. For instance, Hasija et al (2008) explore pay-per-call and pay-per-time contracts in call-centers. In the first type of contracts the vendor earns a fixed fee from the client for each served phone call. In the second type of contracts the vendor is paid per unit of time spent serving customers. In this service setting as in the previous manufacturing ones coordination is achieved via financial transfer payments.

In their path breaking work Su and Zenios (2006) explore the efficiency-equity trade-off in kidney transplantation. In their setting financial transfers are not possible. Instead, they propose a kidney's allocation rule based on the fact that lower-risk patients are willing to spend more time waiting in order to receive organs of higher quality. To the best of our knowledge, this is the first paper using an operational mechanism instead of a financial mechanism for incentive coordination in operations management. Following Su and Zenios, our paper proposes an operational mechanism for incentive coordination. Nevertheless, our work is substantially different from the one of Su and Zenios. First, our research is based on a humanitarian fleet management setting. Second, instead of an allocation rule we propose a capacity (fleet buffer) rule for fleet coordination. Third, in the work of Su and Zenios there are prevailing information rents while in our research we show the existence of natural truth revelation.

Our contribution is twofold. First, we depart from the standard central planning assumption in humanitarian logistics by analyzing a decentralized setting using principal-agent models. Second, this is the first paper using an operational capacity based mechanism for coordination in operations management (table 1).

Stream of literature	Extant literature	This research
Humanitarian	Central planner approach	Decentralized principal-agent approach
Incentives	Financial transfer based mechanisms	Operational capacity based mechanism

Table 1   Paper contribution

3.   Fleet Management in Large International Humanitarian Organizations

With the Headquarter located in the US or Europe, large IHO often implement relief and development Programs in remote areas of developing countries. While relief Programs struggle to estimate their transportation needs in advance, development Programs often have good estimates of their transportation needs. Development Programs use fleets for transportation of staff to visit beneficiaries, transport of materials, and transport of items for distribution to beneficiaries (Pedraza Martinez et al 2010). Although demand is more stable through time than in relief settings, humanitarian development work has some stochasticity - in both arrivals and service. This comes from the mobility of beneficiaries and the unpredictability of operating conditions in terms of weather, road conditions and security. In case of unavailability of vehicles to carry out the visits to beneficiaries the demand tends to accumulate but it rarely disappears. As expressed by a transportation officer in Uganda,

> "*Often I receive more requisitions than the available vehicles I have. When cars are not available, Program staff have to wait.*"

To capture the stochasticity of the system, and the nature of work accumulation when Programs face a fleet shortage, we use a queueing model. The use of queueing models for vehicle fleet management systems is well established in the literature. Queueing models have been used for analyzing police patrol systems (Green 1984, Green and Kolesar 1984a,b, 1989, 2004), fire departments (Kolesar and Blum 1973, Ignall et al 1982), helicopter fire fleets (Bookbinder and Martell 1979), and ambulance fleets (Singer and Donoso 2008). These papers model the needs for transportation using stochastic inter-arrival times.

We represent the Program transportation needs with the Greek letter λ. The transportation needs are measured in rate of visits per unit of time. Program transportation needs are monitored and compiled by the Logistics which make fleet requisitions to the Headquarter. The Headquarter procures the fleet for the Programs.

The sequence of events is as follows: 1) the Headquarter decides the service level of the fleet by choosing a fleet buffer, 2) the Logistics decides the effort of monitoring the Program, 3) the Program states its transportation needs to the Logistics, 4) the Logistics decides whether to monitor the stated needs, takes the monitoring action and requests the fleet to the Headquarter. Table 2 shows the notation used in the modeling.

Decision variables	
γ	Fleet buffer (Headquarter)
p	Monitoring effort (Logistics)
$\hat{\lambda}_i$	Stated transportation needs (Program)
Parameters	
λ_i	True transportation needs, Program type i (Private information)
c	Fleet management cost per vehicle per unit of time
r	Running cost per vehicle per unit of time
w	Cost of delay per visit per unit of time
m	Monitoring cost per unit of time
q	Low type Program probability
Functions	
F	Fleet size
$Q(.)$	Average number of visits in queue per unit of time
$\pi(.)$	Asymptotical approximation to the probability of delay
$\delta_i(.)$	Intended fleet buffer, Program type i
$C_{Head}(.)$	Headquarter cost function
$C_{Prog}(.)$	Program cost function
$C_{Log}(.)$	Logistics cost function
C_{Mec}	Operational mechanism cost function

Table 2   **Notation**

3.1. The Headquarter

The Headquarter has the function of balancing the system costs. These costs include the running cost of the fleet, the fleet management cost in the field, and the cost of delay. The running cost, $r > 0$, is the average running cost per vehicle per unit of time. It includes maintenance, repairs, and fuel costs. The Logistics are accountable for r.

The fleet management cost, $c > 0$, is the average management cost per vehicle per unit of time. Pedraza Martinez et al (2010) find that Programs are accountable for c. Often senior humanitarian staff in the field dedicate a proportion of their time to fleet scheduling and routing, depending on the fleet size. The fleet management cost also includes the cost of vehicle drivers as well as the cost of information systems and spreadsheets to track fleet scheduling and routing in the field. We refer to $c + r$ as the operating cost of the fleet.

The cost of delay, w, is the cost per visit per unit of time. The Programs incur the cost of delay. Although the cost of delay is not a cash cost to the Programs, they incur a disutility associated with delay in carrying out their activities. To simplify our analysis and analytically capture both the disutility of delay and the tensions of this decentralized system we assume a constant marginal cost.

To balance the system costs, i.e. the operating cost and the cost of delay, the Headquarter decides on the fleet size. Denoted by the Greek letter γ, the fleet buffer is the number of vehicles in excess of the minimum number λ/μ. The buffer is required to maintain stability of the system due to inherent variability in inter-arrival and service times. In other words, the fleet buffer is the extra-number of vehicles needed for protecting the system from stochasticity and achieving a determined service level. The optimal fleet buffer can be obtained via optimization methods by balancing the importance of delays in visits with fleet operating costs. We model the system as an M/M/N queue. For analytical tractability we use heavy-traffic approximations under the 'rationalized regime' to represent the average delay in the system. In the rationalized regime, the probability of delay asymptotically converges to a value between 0 and 1. The Programs that are assumed in our system neither carry out emergency response activities nor highly scheduled regular work, but carry out

delay sensitive and stochastically arriving developmental activities. Therefore we believe that the rationalized regime is appropriate for our setting. Borst et al (2004) show that a simple square-root staffing rule is asymptotically optimal for a system operating in the rationalized regime. Hence we assume the square-root staffing rule to determine the fleet size. The square-root rule can be written as $F(\rho) = \rho + \gamma(c, r, w)\sqrt{\rho}$ (Grassmann 1988, Halfin and Whitt 1981, Whitt 1992, Borst et al 2004, Hasija et al 2005). Where $\rho = \frac{\lambda}{\mu}$ is the utilization rate of the fleet as defined in the queueing literature. The effectiveness of the rule increases in the size of the fleet and it has been shown that it is a robust approximation for the optimal system size of systems of 20 or more servers (vehicles).

We model the system as an Erlang-C system following a first-come first-served queueing discipline. This rule was chosen since, as mentioned before, within Programs all the visits have the same priority. For simplicity, we assume $\mu = 1$, which corresponds to measuring time in the scale of mean service times (Whitt 1992). Under this assumption we have:

$$F(\gamma, \lambda) = \lambda + \gamma\sqrt{\lambda} \qquad (1)$$

Following Borst et al (2004) and Hasija et al (2005), for a given fleet buffer γ the average number of visits in the queue is $Q(\gamma, \lambda) = \frac{\pi(\gamma)\lambda}{F\mu - \lambda}$. Using the assumption of $\mu = 1$ and replacing F by its definition in (1), we obtain:

$$Q(\gamma, \lambda) = \frac{\pi(\gamma)\sqrt{\lambda}}{\gamma} \qquad (2)$$

The function $\pi(y)$ is known as the Halfin-Whitt delay function (Halfin and Whitt 1981). It is an asymptotically exact approximation to the probability of delay, $\Pr\{\text{wait} > 0\}$. The value of $\pi(y)$ is:

$$\pi(y) = \left[1 + \frac{y\Phi(y)}{\phi(y)}\right]^{-1}$$

$\Phi(y)$ and $\phi(y)$ are the unit normal cdf and pdf, respectively. The service level of the fleet, S, i.e. the average proportion of visits done without delay is $S = 1 - \pi(y)$.

The Headquarter minimizes the total system cost given by the average cost of delay, $wQ(\gamma, \lambda)$, plus the average operating cost of the fleet, $(c+r)F(\gamma, \lambda)$. The Headquarter problem is:

$$\min_{0<\gamma} \quad C_{Head}(\gamma, \lambda) = wQ(\gamma, \lambda) + (c+r)F(\gamma, \lambda) \tag{3}$$

Replacing (1) and (2) in (3) we can rewrite the Headquarter problem as:

$$\min_{0<\gamma} \quad C_{Head}(\gamma, \lambda) = \frac{w\pi(\gamma)\sqrt{\lambda}}{\gamma} + (c+r)(\lambda + \gamma\sqrt{\lambda}) \tag{4}$$

The cost function (4) is unimodal (Borst et al 2004) and it has a finite and positive minimum value, $\gamma^*(c, r, w)$, which is independent of λ and it only depends on the cost parameters of the system (Hasija et al 2008);

$$\gamma^*(c, r, w) = \arg\min_{0<\gamma} C_{Head}(\gamma, \lambda) \tag{5}$$

is the fleet buffer chosen by the Headquarter.

3.2. The Logistics

The Logistics are accountable for the running cost of the fleet which increases with fleet size. Often the Logistics is composed of national staff located in the capital city of the country of operation. During the planning stage the Logistics compile the Programs' stated transportation needs, denoted by $\hat{\lambda}$. As we will explain later in Section 3.3, the stated transportation needs of the Programs may not be the same as their true transportation needs due to misaligned incentives and private information. The Logistics do not know the true transportation needs, but they have some idea – just not a very accurate one – so we abstract it to two types. We assume that the Logistics has a probabilistic prior belief that:

$$\lambda_i = \begin{cases} \lambda_L, & \text{w.p. } q \\ \lambda_H, & \text{w.p. } 1-q \end{cases}$$

Where L means low transportation needs, H means high needs, $\lambda_L < \lambda_H$ and the space of types is $\Lambda = \{\lambda_L, \lambda_H\}$. This standard assumption helps us to capture the main trade-offs of equity and efficiency keeping the model analytically tractable.

The Logistics can monitor $\hat{\lambda}_i$ before sending the fleet requisition to the Headquarter. To monitor, the Logistics have to carefully check the Program's data on transportation needs. As mentioned before, often Programs have detailed data on fleet use and transportation needs in the field. Typically the data is in printed form, ready for auditing purposes but it is not stored in a way that can be easily accessed by the Logistics at the National level. Due to the lack of trustworthy information systems on transportation needs at the national level, the Logistics often send staff to the field to monitor the Program's estimated workload *in situ*. The Logistics exert a monitoring effort $p \in [0, 1]$ corresponding to the proportion of Programs to monitor. The monitoring cost is $m(p)$, a function of the monitoring effort. By sending staff to the field, the Logistics can get an accurate estimation of the Program's transportation needs. We assume that by monitoring the Logistics can know the true transportation needs of the Program.

The Logistics' objective is to minimize the expected average fleet cost. The average fleet cost is equal to the average running cost plus the monitoring cost. The Logistics problem can be written as:

$$\min_{0 \le p \le 1} \quad E_q[C_{Log}(\gamma, \lambda, p)] = rq[pF(\gamma, \lambda_L) + (1-p)F(\gamma, \hat{\lambda}_L)] \tag{6}$$

$$+ r(1-q)[pF(\gamma, \lambda_H) + (1-p)F(\gamma, \hat{\lambda}_H)] + m(p)$$

3.3. The Program

Ranging from basic health care provision to agriculture and building basic infrastructure, the Program's activities are supported by field vehicle fleets. The transportation needs of the Program relate to the coordination and execution of last mile distribution to beneficiaries. Transportation needs include: 1) transport of relief items in-country and to beneficiaries; 2) transport of staff coordinating or delivering services to beneficiaries, and 3) transport of staff and materials for basic infrastructure Programs (Pedraza Martinez et al 2010).

Each Program has to balance the cost of delay with the cost of managing the fleet in the field. Nevertheless, the Program is not accountable for the running cost of the fleet (Pedraza Martinez

et al 2010). If the Program reports its true transportation needs, the average delay incurred is as in equation (2). Misaligned incentives and private information create an adverse selection issue, and the Program may distort its transportation needs. The i type Program could target an intended fleet buffer δ_i by stating $\hat{\lambda}_i$ as its true transportation needs. Such a distortion would imply an effective buffer of

$$\delta_i(\gamma, \lambda_i, \hat{\lambda}_i) = \frac{\hat{\lambda}_i + \gamma\sqrt{\hat{\lambda}_i} - \lambda_i}{\sqrt{\lambda_i}} \tag{7}$$

Note from (7) that if the low type Program distorts its needs the intended fleet buffer is $\delta_L > \gamma$. On the other hand, if the high type Program distorts its needs the intended fleet buffer is $\delta_H < \gamma$. The expected average queue length is:

$$Q(\delta_i, \lambda_i) = \frac{\pi(\delta_i)\sqrt{\lambda_i}}{\delta_i} \tag{8}$$

The Program's objective is to minimize the service cost $C_{Prog}(\gamma, \lambda_i, \hat{\lambda}_i)$. The i type Program's problem is:

$$\min_{\hat{\lambda}_i \in \Lambda} \quad E_p[C_{Prog}(\gamma, \lambda_i, \hat{\lambda}_i)] = p\left(Q(\gamma, \lambda_i) + cF(\gamma, \lambda_i)\right) + (1-p)\left(Q(x(\hat{\lambda}_i), \lambda_i) + cF(x(\hat{\lambda}_i), \lambda_i)\right) \tag{9}$$

Where

$$x(\hat{\lambda}_i) = \begin{cases} \gamma, & \text{if } \hat{\lambda}_i = \lambda_i \\ \delta_i, & \text{otherwise} \end{cases}$$

The fleet management system is shown in figure 1.

The first thing to note is that there is no penalty, i.e. the Program's stated transportation needs are independent of the Logistics monitoring effort. This result is formalized in the following proposition. All the proofs are included in the appendix.

PROPOSITION 1. *In the current fleet management system the Logistics monitoring effort does not dissuade the Programs from distorting their stated transportation needs.*
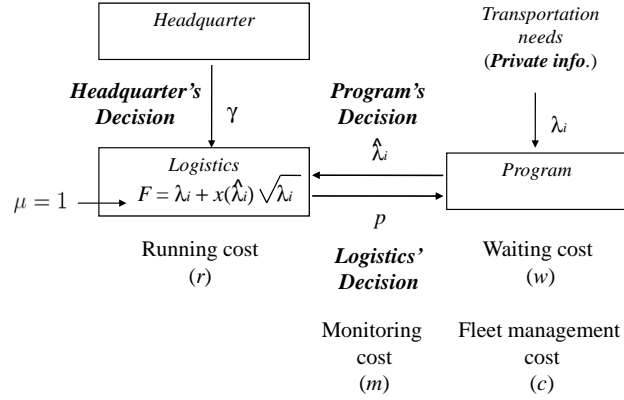
Figure 1   The field vehicle fleet management system

Proposition 1 is consistent with the worries of the Logistics at the national level. As motivated in the Introduction, the Logistics complains about the lack of effective tools to induce the Programs to report their true transportation needs. To find its optimal fleet buffer the Program solves:

$$\min_{0<\gamma} \quad C_{Prog}(\gamma, \lambda_i) = \frac{w\pi(\gamma)\sqrt{\lambda_i}}{\gamma} + c(\lambda_i + \gamma\sqrt{\lambda_i}) \tag{10}$$

The optimal fleet buffer for the Program is:

$$\bar{\gamma}(c, w) = \arg\min_{0<\gamma} C_{Prog}(\gamma, \lambda_i) \tag{11}$$

As in equation (5), the optimal fleet buffer for the Program is independent of its transportation needs. It only depends on the cost parameters c and w. This allows us to state a lemma that clearly explains the misalignment of incentives between the Headquarter and the Program.

LEMMA 1. $\gamma^*(c, r, w) < \bar{\gamma}(c, w)$.

Lemma 1 together with equation (1) imply that the optimal fleet size for the Headquarter is smaller than the optimal fleet size for the Program. This result is consistent with the worries of the Headquarter in terms of oversized fleets and it is also consistent with the worries of the Programs in terms of not having enough vehicles to optimize their service. A graphic representation of the Headquarter cost compared to the Program cost is shown in figure 2.
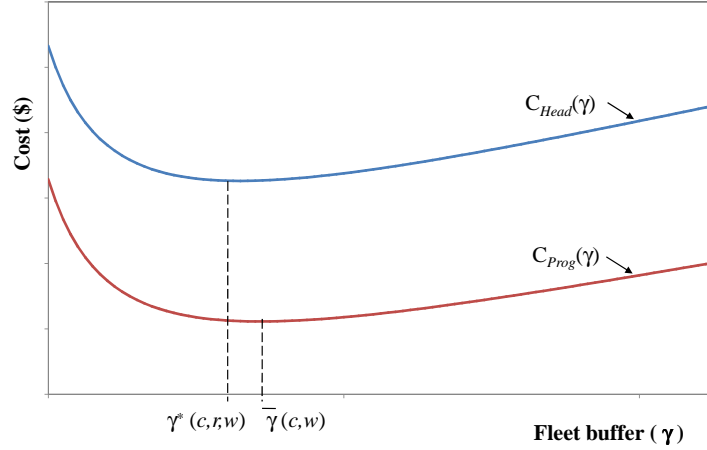
Figure 2 **Headquarter and Program costs as a function of** γ

3.4. Current System Solution

The first observation is that in the current system the high type Program always reveals its true transportation needs. This is formalized in the following proposition.

PROPOSITION 2. *In the current system* $C_{Prog}(\gamma^*, \lambda_H) < C_{Prog}(\delta_H, \lambda_H)$

The intuition behind Proposition 2 comes from the fact that the intended fleet buffer from distortion of transportation needs for the high type Program is lower than the fleet buffer offered by the Headquarter. Additionally, the fleet buffer offered by the Headquarter is lower than the optimal fleet buffer for the Program (Lemma 1). Since the cost function for the Program is unimodal with minimum in $\bar{\gamma}$, it is always cheaper for the high type Program to reveal the truth. Otherwise, the extra-cost of delay would overcome the savings on fleet management.

We are left with the low type Program. The low type Program states its true transportation needs as long as $C_{Prog}(\gamma^*, \lambda_L) < C_{Prog}(\delta_L, \lambda_L)$. This suggests the existence of a threshold for truth telling, formalized in the following proposition.

PROPOSITION 3. *There exists a truth telling threshold* $\hat{\gamma}_L$ *for the low type Program, such that:*

$$C_{Prog}(\hat{\gamma}, \lambda_L) = C_{Prog}(\gamma^*, \lambda_L) \ and \ \hat{\gamma} \neq \gamma^*.$$

Since $\delta_L = f(\gamma^*, \lambda_L, \lambda_H)$, for a fixed value of γ^* the truth telling threshold depends on the ratio between λ_H and λ_L. The result of Proposition 2 and the threshold for truth telling $\hat{\gamma}_L$ allow us to solve the current problem using backward induction. The results are summarized in table 3.

	$\delta_L \in (0, \hat{\gamma}_L]$	$\delta_L \in [\hat{\gamma}_L, \infty]$
$\hat{\lambda}_L$	λ_H	λ_L
$\hat{\lambda}_H$	λ_H	λ_H
p^*	$\min\left[m'^{-1}(rq[F(\delta_L, \lambda_L) - F(\gamma^*, \lambda_L)]), 1\right]$	0
$E[C_{Prog}(\gamma, \lambda_L)]$	$w\left(\frac{p\delta_L \pi(\gamma^*) + (1-p)\gamma^*\pi(\delta_L)}{\gamma^*\delta_L}\right)\sqrt{\lambda_L}$ $+ c(\lambda_L + (p^*\gamma^* + (1-p^*)\delta_L)\sqrt{\lambda_L})$	$wQ(\gamma^*, \lambda_L) + cF(\gamma^*, \lambda_L)$
$E[C_{Prog}(\gamma, \lambda_H)]$	$wQ(\gamma^*, \lambda_H) + cF(\gamma^*, \lambda_H)$	$wQ(\gamma^*, \lambda_H) + cF(\gamma^*, \lambda_H)$
$E[C_{Log}(\gamma, \hat{\lambda}, p)]$	$r[qp^*F(\gamma^*, \lambda_L) + (1 - qp^*)F(\gamma^*, \lambda_H)]$	$r[qF(\gamma^*, \lambda_L) + (1-q)F(\gamma^*, \lambda_H)] + m(p)$

Table 3   **Summary of results for the current system with monitoring**

In the distortion region (second column in table 3), the optimal monitoring effort occurs when the marginal cost of monitoring equals the extra-cost of fleet excess. In the truth telling region (third column in table 3) there is no need for monitoring. This is because the ratio λ_H/λ_L is big enough to guarantee that the Program's savings in cost of delay are overcome by the extra-cost of fleet management. We will come back to the current system with monitoring in our numerical analysis.

4.   An Operational Mechanism Design

In the previous section we showed how our model provides an explanation of the concerns expressed by different staff members of the different parties of the IHO, i.e., the Headquarter, the Logistics, and the Programs. Proposition 1 shows that the lack of penalties associated with the Program's distortion of its transportation needs makes monitoring an ineffective tool of control. Monitoring does not dissuade the low type Program from inflating its needs. Proposition 1 is consistent with the concerns expressed by the Headquarter and the Logistics about the excess of fleet size and the lack of effective monitoring tools during our field interviews. Lemma 1 shows that the Program's optimal fleet buffer is greater than the headquarter optimal fleet buffer. This explains the Program's concerns of insufficient fleet size. Proposition 2 shows that the high type Program has no incentive

to distort its true transportation needs. But the low type Program has incentives to distort its needs as long as the intended fleet buffer is below the truth telling threshold defined in Proposition 3.

Since financial transfer payments are not implementable in this setting, truth revelation proves to be challenging. In this section we propose a novel mechanism design for truth revelation based on an operational lever, the fleet buffer offered by the Headquarter.

The Headquarter (Principal), commits to a mechanism that allocates the Program (Agent) type i an outcome F, i.e. a fleet size, as a function of the reported type. This is operationalized by the Headquarter committing to an offered fleet buffer, γ_i, for the stated transportation need of the Program equal to λ_i. The Program reports a type profile $\hat{\lambda}_i \in \Lambda$ and the mechanism is executed. The Headquarter's objective is to minimize the system cost for a given distribution of Program types:

$$\min_{\gamma_L, \gamma_H} E[C_{Mec}] = q\left[wQ(\gamma_L, \lambda_L) + (c+r)F(\gamma_L, \lambda_L)\right] + (1-q)\left[wQ(\gamma_H, \lambda_H) + (c+r)F(\gamma_H, \lambda_H)\right] \quad (12)$$

The incentive compatibility (IC) constraints consist of offering each Program type $i = \{L, H\}$ a fleet buffer γ_i to guarantee that $C_{Prog}(\gamma_i, \lambda_i) \leq C_{Prog}(\delta_i, \lambda_i)$. Hence, for each Program type the (IC) constraint would be:

$$wQ(\gamma_i, \lambda_i) + cF(\gamma_i, \lambda_i) \leq wQ(\delta_i, \lambda_i) + cF(\delta_i, \lambda_i) \quad (13)$$

The left hand side of constraint (13) is the i type Program cost with a fleet buffer γ_i. The right hand side of (13) represents the cost of the i type Program when distorting its needs. The intended fleet buffer δ_i is defined by (7).

In the system we are modeling the Program does not have an outside option since the fleet is procured by the Headquarter. Hence, the individual rationality (IR) constraints are defined by the fact that the Headquarter must offer each Program type a fleet buffer big enough to protect the

system from stochasticity, i.e. $0 < \gamma_i$. Combining objective (12) with the *IC* constraints (13) and the *IR* stability constraints, the mechanism can be stated as:

$$\min_{\gamma_L, \gamma_H} E[C_{Mec}] = q\left[wQ(\gamma_L, \lambda_L) + (c+r)F(\gamma_L, \lambda_L)\right] + (1-q)\left[wQ(\gamma_H, \lambda_H) + (c+r)F(\gamma_H, \lambda_H)\right]$$

$$S.T. \tag{14}$$

$$(IC_L): \quad wQ(\gamma_L, \lambda_L) + cF(\gamma_L, \lambda_L) \leq wQ(\delta_L, \lambda_L) + cF(\delta_L, \lambda_L)$$

$$(IC_H): \quad wQ(\gamma_H, \lambda_H) + cF(\gamma_H, \lambda_H) \leq wQ(\delta_H, \lambda_H) + cF(\delta_H, \lambda_H)$$

$$(IR_L): \quad 0 < \gamma_L$$

$$(IR_H): \quad 0 < \gamma_H$$

In the proposed mechanism (14) the Logistics is not a strategic decision maker. Instead, the Logistics focus on applying fleet management policies and procedures designed by the Headquarter.

There are two ways of achieving truth revelation via the operational mechanism: induced and natural. Natural truth revelation follows directly from Proposition 3. We will now show that when conditions in Proposition 3 are not met, there exist fleet buffers γ_L and γ_H such that both Program types have incentives to reveal their true transportation needs. In the case of induced truth revelation we find that the *IC* constraint of the low type Program will always be binding. Let

$$\tilde{\gamma}_L = \{\gamma \neq \gamma_L : C_{Prog}(\gamma) = C_{Prog}(\gamma_L)\} \tag{15}$$

There exist two ways of making the low type Program's *IC* constraint tight. The first way is by forcing $\gamma_L = \delta_L$. The second way of making low type Program's *IC* constraint tight is by choosing γ_H such that $\delta_L = \tilde{\gamma}_L$.

PROPOSITION 4. *There exist two regions for induced truth revelation via the operational mechanism as follows:*

1) R1: Equal fleet size region. $\gamma_H < \gamma^* < \gamma_L$ such that $F_L = F_H$

2) R2: Different fleet size region. $\gamma^* < \gamma_L$, $\gamma^* < \gamma_H$ *such that* $F_L < F_H$ *and the regions for induced truth telling are separated by the threshold* T_1.

The two regions characterized by Proposition 4 can be depicted in the space defined by the transportation needs ratio and probability types (figure 3). In *R1* of Proposition 4 both *IC* constraints in mechanism (14) are binding. The Headquarter offers $\delta_H = \gamma_H < \gamma^* < \gamma_L = \gamma_L$ such that both Program types receive the same fleet size, $F_L = F_H$. This region exists for values of λ_H/λ_L close to 1 and limited by T_1, a threshold implicitly defined by the cost parameters of the system, c, r and w, the lambda ratio, λ_H/λ_L and the low type probability, q. When parameters fall in *R1*, truth revelation is achieved by making the low type Program indifferent between reporting its true needs and inflating these needs, i.e. by making the low type *IC* constraint tight via $\gamma_L = \delta_L$. The extra-cost for low type's truth revelation is mitigated by reducing δ_L via the decrease of γ_H, such that $\gamma_H = \delta_H$. In our numerical experiments we find that the cost mitigation is mediated by q, the low type probability. Increasing q after the parameter dependent threshold T_1 reduces the width of *R1*.

In *R2* of Proposition 4 only the *IC* constraint for the low type Program in (14) is binding. Both γ_L and γ_H are greater than γ^*, which means that the Headquarter gives incentives to both Program types. For low values of q the Headquarter keeps γ_H closer to γ^* by making $\gamma^* < \gamma_H < \gamma_L$. This decision is due to the likelihood of having a larger proportion of high type Programs. This way the Headquarter avoids a big increase in the high type Program cost while making the low type Program *IC* constraint tight, via $\delta_L = \tilde{\gamma}_L$. For large values of q the increase in the likelihood of having low type Programs switches the order of incentives to $\gamma^* < \gamma_L < \gamma_H$. The closer γ_L is to γ^* the lower the cost for the headquarter. As mentioned earlier, we find that as q increases, the region *R2* becomes more favorable than the region *R1*. The intuition is straightforward. In *R2*, $F_L < F_H$ while in *R1*, $F_L = F_H$. As q increases, the probability of a Program to be of the low type increases. Hence $F_L < F_H$ becomes a more cost effective mechanism than $F_L = F_H$.

The third region in figure 3, *R3*, is the region of natural truth revelation. Under some parameter
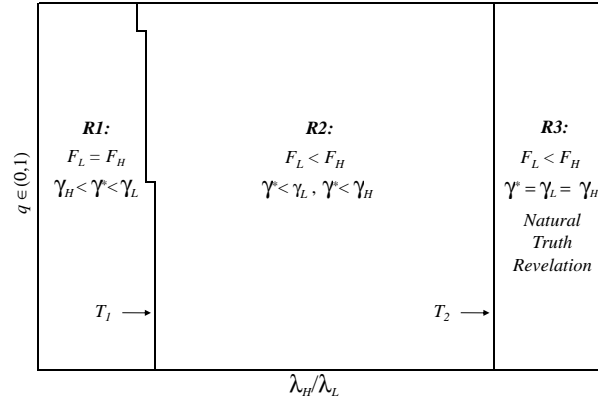
Figure 3 **Truth telling regions in the transportation needs ratio and probability type** $(\gamma_H/\gamma_L, q)$ **space**

values neither of the Program types has incentives to distort their transportation needs, i.e. there exists a natural truth telling region.

COROLLARY 1. *There exists a "natural truth telling" region for the operational mechanism. R3, the natural truth telling region is defined by* $\gamma^* = \gamma_L = \gamma_H$ *such that* $F_L < F_H$.

In *R3*, characterized in Corollary 1 both *IC* constraints are loose. Truth revelation is achieved without the need of incentives. With parameter values falling in this region the low type Program does not have incentives for distorting its needs, i.e.$\tilde{\gamma} < \delta_L$. If the low type Program inflates its transportation needs, then it gets a fleet big enough to guarantee that the cost of management will overcome the savings in delay. On the other hand, the high type Program does not distort its needs because it would receive a fleet too small for its needs. In this case the cost of delay would overcome the savings in fleet management. It is important to remark that in figure (3) the threshold for natural truth revelation in the mechanism, T_2 equals the truth telling threshold in the current system, $\hat{\gamma}$, defined in Proposition 3. This follows from the fact that the Headquarter's cost function in both formulations is equivalent. It is also interesting that T_2 is independent of q. It depends on the cost parameters c, r, w, and the λ_H/λ_L ratio.

Another interesting insight is that the high type Program does not distort its transportation needs. This is formalized in the following proposition.

PROPOSITION 5. *Under the mechanism the high type Program always reveals its true transportation needs.*

In *R1* characterized in Proposition 4 the high type Program is indifferent between revealing its true transportation needs and distorting. This is because $\gamma_H = \delta_H$. In *R2* also characterized in Proposition 4 the high type Program is better of by revealing its true transportation needs since condition $\delta_H < \gamma_H < \bar{\gamma}$ implies $C_{Prog}(\gamma_H) < C_{Prog}(\delta_H)$. Finally, in *R3*, characterized in Corollary 1, truth revelation is achieved without the need of incentives.

By being flexible in choosing fleet buffers γ_L and γ_H instead of a unique γ^*, the Headquarter can create operational incentives for truth revelation. These incentives increase the system's efficiency without the need of monitoring the Program's reported needs.

COROLLARY 2. *The operational mechanism is optimal for the headquarter compared to the current system without monitoring.*

In fact, the operational mechanism allows the Headquarter to achieve significant cost savings compared to the current system. These savings as well as the sensitivity analysis of the mechanism regarding the waiting cost will be illustrated in the next section.

5. Numerical study

This section presents a numerical study that complements the analytical insights presented in the characterization of the operational mechanism. The base case uses weekly planning for a time horizon of 52 weeks. The running cost and the fleet management cost of the fleet are calculated following the research by Pedraza Martinez and Van Wassenhove (2010) on vehicle replacement in a humanitarian setting. The running cost per vehicle is established at $17,000 per year. This is equivalent to $269,23 per week. It includes maintenance, repairs and fuel. The fleet management cost is established to be 15% of the running cost. It includes the time of staff coordinating fleet management, and the salary of vehicle drivers, which depend on the Program (Pedraza Martinez et al 2010). The normalized demand rate for the low type Program is $\lambda_L = 60$ visits per week. This

is equivalent to a fleet of 60 vehicles with a utilization rate of 100%. For the current system with monitoring we assume that the monitoring cost equals the fleet management cost. In the numerical examples we use a convex monitoring cost function $m(p) = m[qF(\gamma, \lambda_L) + (1-q)F(\gamma, \lambda_H)]p^2$. This signifies that the monitoring cost is proportional to the expected fleet size under the current system. Figure 4 shows a comparison of cost for: 1) centralized benchmark; 2) current system without monitoring; 3) current system with monitoring; 4) operational mechanism.
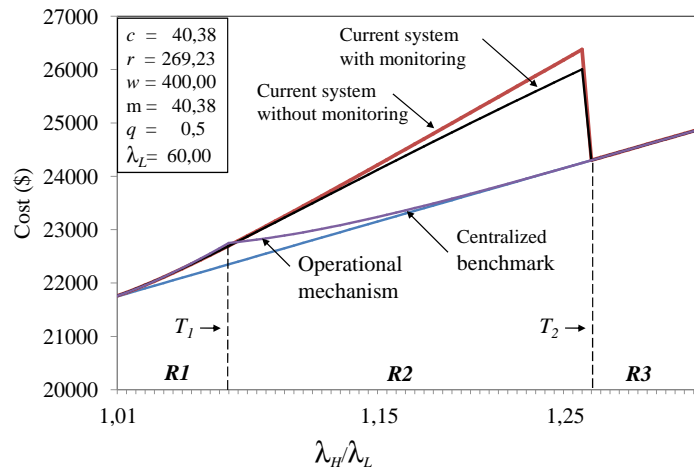

Figure 4   System cost comparison

The cost of the current system without monitoring suffers from a fleet excess caused by the inflation of transportation needs coming from the low type Program. This inflation holds for parameter values below the threshold for "natural truth revelation", T_2. Note that the cost of the current system without monitoring is an upper bound for the cost of the system with monitoring. The greater the monitoring cost, the closer the cost of the current system with monitoring to the upper bound. The centralized benchmark is a lower bound for the current system with monitoring. The lower the monitoring cost, the closer the current system with monitoring will be to the centralized benchmark.

Also note that in *R1* the mechanism does not produce significant savings compared to the current system. In this region the Headquarter offers the same fleet size to both Programs. This strategy makes the low type Program indifferent between revealing the truth and inflating its needs. In *R2*

the efficacy of the mechanism in terms of cost reduction increases compared to the current system. By being flexible in assigning different fleet buffers to both Program types the Headquarter can diminish the fleet size inflation via induced truth revelation.

The fleet buffers can be plotted in the $(\lambda_H/\lambda_L, q, \gamma)$ space (figure 5). the horizontal axis represents the transportation needs ratio λ_H/λ_L. The depth axis represents the low type Program probability q. The vertical axis represents the fleet buffer for the i type Program. Figure 5 is complemented with figure 6 showing the projection of γ_L and γ_H in the $(\lambda_H/\lambda_L, \gamma)$ space, and figure 7 showing the projection of γ_L and γ_H in the (q, γ) space. First, note that the Headquarter gives incentives to the low type Program in the induced truth region $R1$ by choosing $\gamma^* < \gamma_L$ (figures 6a and 7a). On the other hand, to mitigate the extra-cost of this strategy, the Headquarter simultaneously decreases the fleet buffer for the high type Program by choosing $\gamma_H < \gamma^*$ (figure 6b and $\lambda_H/\lambda_L = 1,01$ in figure 7b). This way the Headquarter forces the indifference conditions $\gamma_H = \delta_H < \gamma^* < \gamma_L = \delta_L$. For low values of q in $R1$ both γ_L and γ_H are very close to γ^*. When q increases, the Headquarter offers a lower γ_H, which pushes down δ_L. The Headquarter makes the low type Program indifferent between revealing the truth and inflating its needs by offering $\gamma_L = \delta_L$. The extra cost of delay for the high type Program is compensated via reduction of operating cost for the fleet of the low type Program.

In $R2$ the Headquarter's strategy is different. For low values of q the Headquarter gives incentives to both Program types by choosing $\gamma^* < \gamma_H < \gamma_L < \bar{\gamma} < \delta_L = \tilde{\gamma}$. To keep the system cost balanced, the Headquarter offers the low type Program a fleet buffer greater than γ^*, but decreasing in the proportion of low type Programs, q, (figure 7a). In fact, for large values of q it is more efficient for the Headquarter to offer $\gamma^* < \gamma_L < \gamma_H < \bar{\gamma} < \delta_L = \tilde{\gamma}$. Nevertheless, decreasing γ_L increases $\tilde{\gamma}$, such that to have the indifference condition $\delta_L = \tilde{\gamma}$, the Headquarter must offer the high type Program a bigger fleet buffer, up to $\gamma_H = 1,7364$ when $q = 0,975$ (figure 7b).

To summarize, in $R1$ the Headquarter offers the same fleet size to both Programs by offering the low type Program $\gamma^* < \gamma_L$. Simultaneously, the high type is offered a fleet buffer $\gamma_H < \gamma^*$. This way the Headquarter induces truth revelation for the low type Program while keeping the high

(a) Low type Program
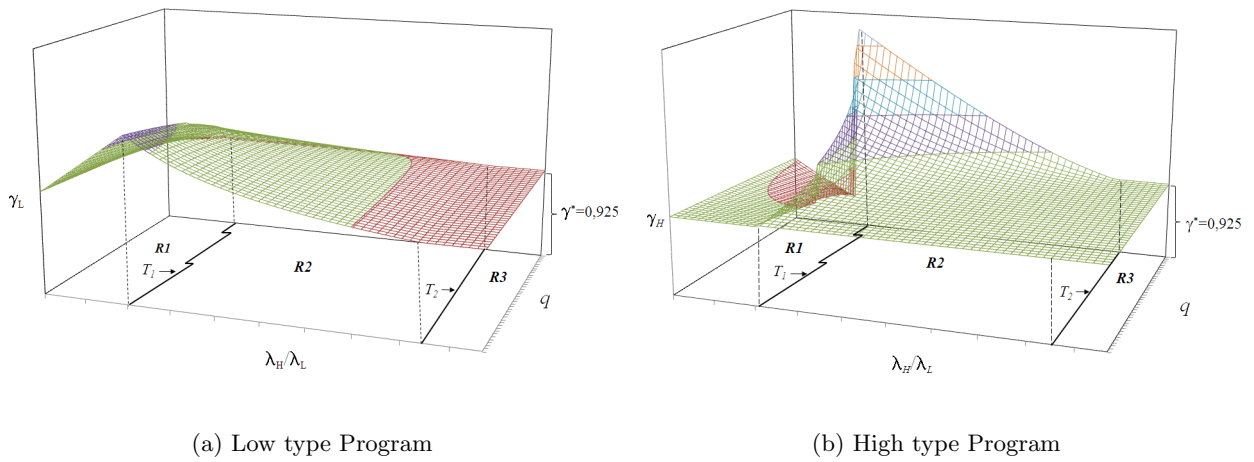
(b) High type Program

Figure 5 **Fleet buffers in the** $(\lambda_H/\lambda_L, q, \gamma)$ **space**

(a) Low type Program
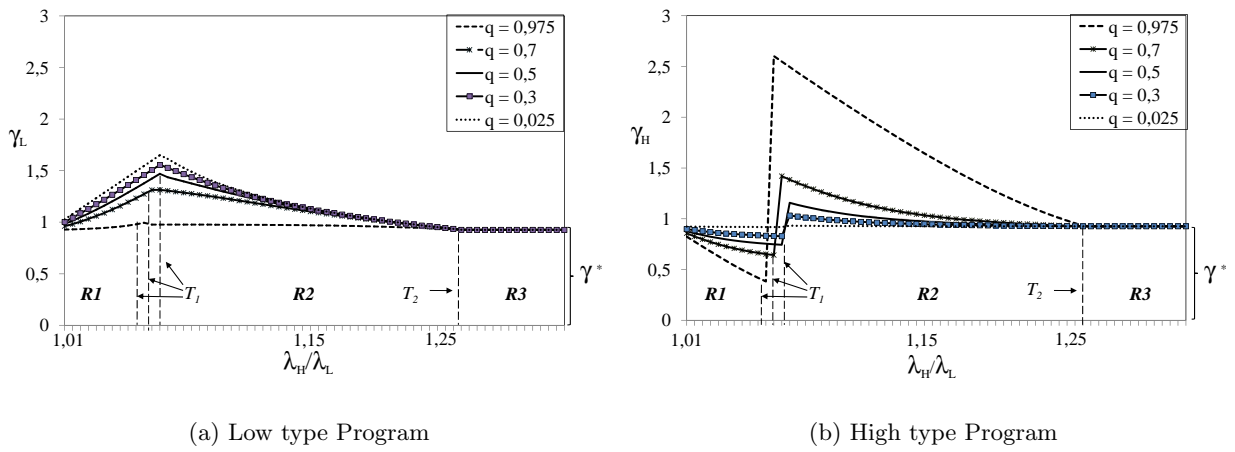
(b) High type Program

Figure 6 **Fleet buffers as function of** λ_H/λ_L

(a) Low Type Program
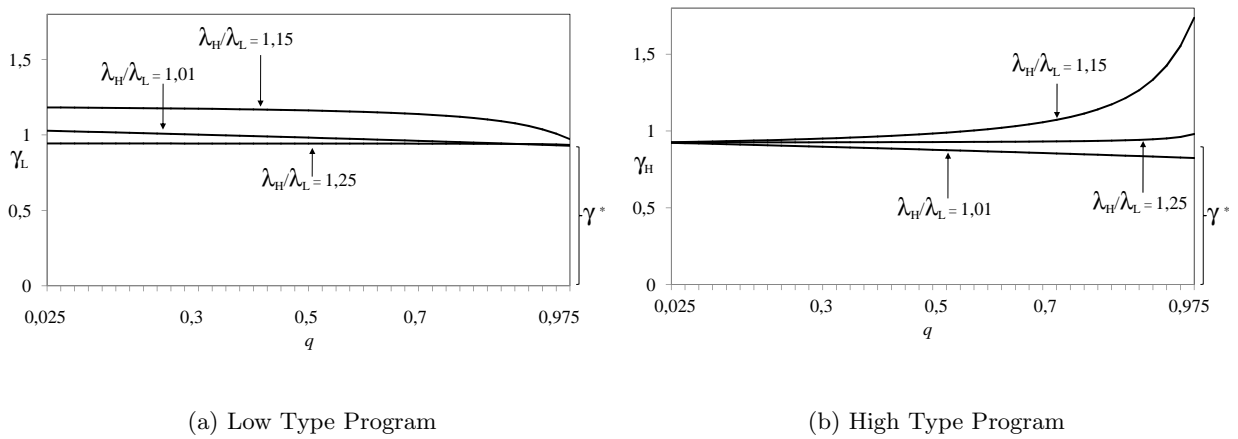
(b) High Type Program

Figure 7 **Fleet buffers as function of** q

type Program indifferent. In $R2$ the Headquarter makes the low type Program indifferent between revealing its true needs and inflating them by offering $\gamma^* < \gamma_L$. But the system cost is controlled by choosing γ_L according to the proportion of low type Programs. The higher the proportion of low type Programs, the closer γ_L is to γ^*. In $R3$ there is no need for incentives since truth revelation is achieved naturally.

Next we perform a sensitivity analysis with respect to w. Since we do not have data on the possible cost of delay, we study the sensitivity of the system to changes in this parameter by choosing $w = \{100; 200; 400; 600; 800\}$. We also use $q = \{0,025; 0,3; 0,5; 0,7; 0,975\}$ and lambda ratios $\lambda_H/\lambda_L = \{1,01; 1,15; 1,25\}$. The sensitivity analysis summarizes the changes in the Headquarter's optimal fleet buffer, γ^* and the Program's optimal fleet buffer, $\bar{\gamma}$ (table 4).

	$\lambda_H/\lambda_L = 1,01$					$\lambda_H/\lambda_L = 1,15$					$\lambda_H/\lambda_L = 1,25$				
w =	100	200	400	600	800	100	200	400	600	800	100	200	400	600	800
						Fleet Buffer Measures[†]									
γ^*	0,529	0,710	0,925	1,064	1,166	0,529	0,710	0,925	1,064	1,166	0,529	0,710	0,925	1,064	1,166
$\bar{\gamma}$	1,246	1,468	1,701	1,840	1,939	1,246	1,468	1,701	1,840	1,939	1,246	1,468	1,701	1,840	1,939
$T_1^{\ddagger}$	1,055	1,065	1,070	1,070	1,070	1,055	1,065	1,070	1,070	1,070	1,055	1,065	1,070	1,070	1,070
T_2	1,220	1,250	1,265	1,265	1,270	1,220	1,250	1,265	1,265	1,270	1,220	1,250	1,265	1,265	1,270
q						**Low Type Program Fleet Buffer,** γ_L									
0,025	0,630	0,811	1,028	1,167	1,270	0,673	0,924	1,182	1,331	1,434	0,529	0,710	0,943	1,088	1,191
0,300	0,607	0,787	1,002	1,141	1,244	0,670	0,918	1,174	1,322	1,425	0,529	0,710	0,943	1,088	1,191
0,500	0,588	0,767	0,982	1,121	1,223	0,665	0,909	1,162	1,310	1,413	0,529	0,710	0,943	1,087	1,191
0,700	0,567	0,746	0,961	1,099	1,202	0,655	0,890	1,138	1,284	1,388	0,529	0,710	0,942	1,087	1,190
0,975	0,533	0,713	0,928	1,067	1,170	0,560	0,750	0,972	1,114	1,217	0,529	0,710	0,934	1,075	1,178
q						**High Type Program Fleet Buffer,** γ_H									
0,025	0,527	0,708	0,923	1,062	1,164	0,530	0,711	0,927	1,066	1,168	0,529	0,710	0,925	1,064	1,167
0,300	0,504	0,683	0,898	1,036	1,138	0,541	0,729	0,950	1,091	1,193	0,529	0,710	0,926	1,066	1,168
0,500	0,486	0,664	0,878	1,016	1,118	0,557	0,755	0,983	1,125	1,228	0,529	0,710	0,928	1,068	1,170
0,700	0,465	0,643	0,856	0,994	1,096	0,594	0,815	1,057	1,201	1,304	0,529	0,710	0,931	1,072	1,175
0,975	0,431	0,610	0,824	0,962	1,064	1,018	1,404	1,736	1,901	2,006	0,529	0,710	0,980	1,138	1,243

[†] The fleet buffer measures remain constant to changes in λ_H/λ_L
[‡] $q = 0,5$

Table 4   Sensitivity of results to changes in w

A decrease in the cost of delay, w, causes a decrease in both the optimal fleet buffer for the Headquarter, γ^* and in the optimal fleet buffer for the Program $\bar{\gamma}$. The lower the cost of delay, the lower the relative weight of the equity component of the objective function for the system. On the other hand, an increase in w increases γ^* and $\bar{\gamma}$. Nevertheless, these changes in the optimal fleet buffer are not linear in w. Decreasing w by 50% starting at the $\lambda_H/\lambda_L = 1,15$ case decreases

γ^* by 23,24% while increasing w by 50% only increases γ^* by 15,02%. Note that decreasing w expands *R3*, the natural truth telling region. Lower w implies that the fleet management cost for the Program is relatively higher. Hence the Program has lower incentives to distort transportation needs to secure a bigger fleet. On the other hand, increasing w slightly decreases *R3*.

We also summarize the changes in the natural truth telling threshold for the current system and for the operational mechanism. Our sensitivity analysis includes the cost of: 1) centralized benchmark, 2) current system without monitoring, 3) current system with monitoring, 4) mechanism. Results are presented in table 5.

	$\lambda_H/\lambda_L = 1,01$					$\lambda_H/\lambda_L = 1,15$					$\lambda_H/\lambda_L = 1,25$				
w =	100	200	400	600	800	100	200	400	600	800	100	200	400	600	800
q						**Centralized Benchmark Cost**									
0,025	20743	21268	21844	22196	22448	23409	23968	24582	24956	25225	25308	25890	26530	26920	27200
0,300	20689	21213	21789	22140	22392	22603	23152	23754	24122	24386	23967	24532	25153	25532	25803
0,500	20650	21174	21749	22099	22351	22017	22558	23153	23515	23775	22991	23544	24152	24522	24788
0,700	20611	21134	21708	22059	22310	21431	21965	22551	22908	23165	22016	22556	23150	23512	23772
0,975	20557	21080	21653	22003	22254	20626	21149	21723	22074	22325	20674	21198	21773	22124	22376
q						**Operational Mechanism Cost**									
0,025	20744	21269	21845	22196	22449	23411	23971	24586	24961	25230	25308	25890	26530	26920	27200
0,300	20699	21222	21797	22147	22399	22628	23194	23808	24177	24440	23967	24532	25153	25532	25804
0,500	20664	21185	21759	22109	22360	22057	22627	23238	23603	23862	22991	23544	24152	24523	24789
0,700	20623	21144	21717	22067	22318	21483	22053	22659	23020	23276	22016	22556	23151	23514	23774
0,975	20559	21081	21654	22004	22255	20647	21181	21762	22114	22366	20674	21198	21774	22125	22377
q						**Current System With Monitoring Cost**									
0,025	20744	21269	21845	22196	22449	23463	24021	24635	25009	25278	25308	25890	26630	27021	27302
0,300	20698	21221	21796	22146	22398	23224	23760	24358	24728	24995	23967	24532	26253	26638	26917
0,500	20665	21187	21760	22111	22362	23008	23531	24120	24487	24752	22991	23544	25841	26224	26501
0,700	20632	21152	21725	22075	22326	22754	23266	23849	24212	24477	22016	22556	25297	25678	25954
0,975	20587	21105	21677	22026	22277	22334	22839	23415	23776	24039	20674	21198	24299	24678	24953
q						**Current System Without Monitoring Cost**									
0,025	20745	21270	21846	22197	22450	23465	24022	24636	25010	25280	25308	25890	26632	27023	27303
0,300	20699	21222	21797	22147	22399	23263	23795	24391	24761	25028	23967	24532	26370	26756	27035
0,500	20666	21187	21761	22111	22363	23116	23629	24214	24579	24845	22991	23544	26180	26561	26839
0,700	20632	21152	21725	22075	22326	22969	23464	24036	24397	24662	22016	22556	25990	26367	26644
0,975	20586	21105	21676	22025	22276	22767	23236	23792	24147	24410	20674	21198	25728	26100	26375

Table 5   Sensitivity of system costs to changes in w

Note that the effectiveness of the mechanism compared to the current system without monitoring is increasing in w, λ_H/λ_L and q, as long as parameter values fall into the induced truth telling regions, *R1* and *R2*. While the base case of $w = 400$, $\lambda_H/\lambda_L = 1,15$ and $q = 0,5$ leads to savings of 4,03%, the case of $w = 800$, $\lambda_H/\lambda_L = 1,15$, $q = 0,975$ produces savings of 15,16% compared to the current system without monitoring.

This section presented a cost comparison between the different regimes analyzed in the paper. It also illustrated graphically the Headquarter's strategies to achieve truth revelation using the operational mechanism. These strategies vary depending on parameter values. While in $R1$ the Headquarter offers the same fleet size to both Program types, in $R2$ the strategy consists of offering incentives to both Program types and keeping the buffer for the Program with the biggest expected proportion closer to γ^*. Finally, this section included a sensitivity analysis with respect to w. The effectiveness of the mechanism compared to the current system increases with w as long as the parameters fall into the induced truth regions. The next section presents the conclusions and possible avenues for future research.

6.   Conclusions and Further Research

This paper studies a decentralized three party fleet management system in a humanitarian setting. Located in the field, Programs are service oriented and have private information on their transportation needs. Transportation needs are fulfilled using vehicle fleets. Located in the main cities in the countries of operation, the Logistics main objective is efficiency via minimization of the running cost of the fleet. Logistics can monitor the Program's stated transportation needs to avoid this inflation. Often located in the US or Europe the Headquarter must balance the cost of delay and the operating cost of the fleet. Differences in objectives, distant geographic locations and the private information of Programs about their transportation needs give rise to an adverse selection problem. The headquarter is concerned about the excess of fleet size. The Logistics concern relates to the lack of effectiveness of monitoring tools. Finally, the Programs are concerned about the lack of vehicles to respond to their transportation needs.

We find that the concerns of the three parties in the system are rational from each party's perspective. In the current fleet management system the Logistics monitoring effort does not dissuade the low type Program from inflating transportation needs. This is because of the impossibility of punishing the distorted needs stated by the Program. We also find that the optimal fleet buffer offered by the Headquarter is lower than the optimal fleet buffer intended by the Program. Hence,

the lack of effectiveness of monitoring tools added to the incentive of the low type Program to inflate transportation needs produce a fleet excess that justifies the Headquarter concerns. Finally, the Program's concerns about not having enough vehicles to respond to its needs are explained by the fact that its optimal fleet buffer is greater than the one offered by the Headquarter.

Nevertheless, we find that for appropriate parameter combinations the current system has a "natural" truth telling threshold in which the centralized benchmark solution can be achieved. In the current fleet management system the high type Program always reveals its true transportation needs. Otherwise, the high type Program would receive a lower fleet buffer compared to the one offered by the Headquarter, increasing even more this Program's costs of delay. The threshold for natural truth telling arises when the extra fleet management cost for the fleet excess intended by the low type Program dominates the savings from the reduction in the cost of delay.

The coordination of incentives is particularly challenging since financial transfer payments are not a viable way to induce Program's truth revelation in this humanitarian setting. We use this humanitarian setting to show the strategic importance of operational design, beyond the objective to achieve tactical efficiency in the humanitarian sector.

We propose a novel operational capacity based mechanism to coordinate incentives in this system. In this mechanism the Headquarter offers different fleet buffers to the different Program types. Additionally, the monitoring role of logistics is suppressed and the Logistics focuses on the roles of implementing the policies and procedures defined by the Headquarter. We show the existence of three mutually exclusive and collectively exhaustive regions for truth revelation under the proposed mechanism. *R1* is called the equal fleet size region. *R2* is called the different fleet size region. Finally, *R3* is called the natural truth revelation region.

In the equal fleet size region the Headquarter achieves truth revelation by offering both Program types the same fleet size. This strategy makes the low type Program indifferent between inflating its transportation needs and revealing the truth. In the different fleet size region both Program types are offered fleet buffers greater than optimal fleet buffer for the current system. When small proportions of low type Programs are expected, the low type Program is offered a fleet buffer such

that this Program's cost when inflating its needs equals the cost of revealing its true needs. When the low type probability is high enough, it is cheaper for the Headquarter to incentivize the high type Program. By offering the high type a bigger fleet buffer, the Headquarter increases the low type Program's intended fleet buffer enough to make this Program indifferent between revealing its needs and inflating them. Finally, under some parameter combination which is independent of the low type probability, the system reaches a natural truth telling region. In this region, as in the current system, the extra-cost of fleet management deters the low type Program from inflating its transportation needs. Additionally, we show that under the mechanism the high type Program always reveals his true transportation needs. This result is similar to the one we got for the current system.

Our numerical experiments complement the analysis by showing the behavior of the system due to changes in the cost of delay. As expected, a decrease in the cost of delay decreases the optimal fleet buffers for the system. An increase in the cost of delay increases the fleet buffers for the system. It is interesting to remark that the change in the optimal fleet buffers is not linear to changes in the cost of delay. Equally, the thresholds defining the regions for truth telling are much more sensitive to decreases in the cost of delay than they are to the increases in that cost. This is because the decrease of the cost of delay makes the Program's cost function flatter rapidly, increasing the size of the natural truth revelation region.

Our paper opens a new stream of research on coordination via operational capacity based mechanism design. This is suitable to settings in which financial transfer payments cannot be used to coordinate incentives in decentralized systems.

Appendix. Proofs

Proof of Proposition 1: A Program of type i would report the true transportation needs if in (9) we have $wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i) \leq p(wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i)) + (1-p)(wQ(\delta_i, \lambda_i) + cF(\delta_i, \lambda_i))$. Note that $wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i) = p(wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i)) + (1-p)(wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i))$. Using this fact we get the truth telling condition $wQ(\gamma, \lambda_i) + cF(\gamma, \lambda_i) \leq wQ(\delta_i, \lambda_i) + cF(\delta_i, \lambda_i)$. The proof follows from the fact that the condition for truth telling does not depend on p, the Logistics monitoring effort. □

Proof of Lemma 1: Note that:

$$\frac{\partial C_{Head}(\gamma,\lambda)}{\partial \gamma} = \left[\frac{w\pi'(\gamma)}{\gamma} - \frac{w\pi(\gamma)}{\gamma^2} + c + r \right] \sqrt{\lambda} \tag{16}$$

and:

$$\frac{\partial C_{Prog}(\gamma,\lambda)}{\partial \gamma} = \left[\frac{w\pi'(\gamma)}{\gamma} - \frac{w\pi(\gamma)}{\gamma^2} + c \right] \sqrt{\lambda} \tag{17}$$

Hence

$$\frac{\partial C_{Prog}(\gamma,\lambda)}{\partial \gamma} = \frac{\partial C_{Head}(\gamma,\lambda)}{\partial \gamma} - r\sqrt{\lambda}$$

$$\frac{\partial C_{Prog}(\gamma,\lambda)}{\partial \gamma}\Big|_{\gamma=\gamma^*(c,r,w)} = \frac{\partial C_{Head}(\gamma,\lambda)}{\partial \gamma}\Big|_{\gamma=\gamma^*(c,r,w)} - r\sqrt{\lambda}$$

$$\frac{\partial C_{Prog}(\gamma,\lambda)}{\partial \gamma}\Big|_{\gamma=\gamma^*(c,r,w)} = 0 - r\sqrt{\lambda} < 0$$

Because $C_{Prog}(\gamma,\lambda)$ is unimodal with a finite minimum, $C_{Prog}(c,w)$ decreases for values of γ such that $\gamma^*(c,r,w) < \gamma$, reaching its minimum at $\gamma = \bar{\gamma}(c,w)$. This implies $\gamma^*(c,r,w) < \bar{\gamma}(c,w)$. $\square$

Proof of Proposition 2: Given the Headquarter solution γ^*, we have $\delta_H = \frac{\lambda_L + \gamma^*\sqrt{\lambda_L} - \lambda_H}{\sqrt{\lambda_H}}$. By construction, $\delta_H < \gamma^*$. Since $\frac{\partial C_{Prog}(\gamma,\lambda_H)}{\partial \gamma}\Big|_{\gamma=\gamma^*} < 0$, and due to the fact that the cost function $C_{Prog}(\gamma,\lambda_H)$ is unimodal with minimum at $\bar{\gamma}$, and given that from lemma 1 we know that $\gamma^* < \bar{\gamma}$, it follows that $C_{Prog}(\gamma^*,\lambda_H) < C_{Prog}(\delta_H,\lambda_H)$. $\square$

Proof of Proposition 3: Note that $\bar{\gamma} < \hat{\gamma}_L$. If $\delta_L < \hat{\gamma}_L$, then $C_{Prog}(\delta_L,\lambda_L) < C_{Prog}(\gamma^*,\lambda_L)$. If $\delta_L = \hat{\gamma}_L$, then $C_{Prog}(\delta_L,\lambda_L) = C_{Prog}(\gamma^*,\lambda_L)$. Finally, if $\hat{\gamma}_L < \delta_L$, then $C_{Prog}(\gamma^*,\lambda_L) < C_{Prog}(\delta_L,\lambda_L)$. Hence, $\hat{\gamma}_L$ defines the truth telling region for the low type Program. $\square$

Proof of Proposition 4: First, by replacing (1) and (2) in (14) we rewrite the mechanism in extended form as:

$$\min_{\gamma_L,\gamma_H} E[C_{Mec}] = q\left[\frac{w\pi(\gamma_L)\sqrt{\lambda_L}}{\gamma_L} + (c+r)(\lambda_L + \gamma_L\sqrt{\lambda_L}) \right] \tag{18}$$
$$+ (1-q)\left[\frac{w\pi(\gamma_H)\sqrt{\lambda_H}}{\gamma_H} + (c+r)(\lambda_H + \gamma_H\sqrt{\lambda_H}) \right]$$

$\quad$ S.T.

$$(IC_L): \quad \frac{w\pi(\gamma_L)\sqrt{\lambda_L}}{\gamma_L} + c(\lambda_L + \gamma_L\sqrt{\lambda_L}) \leq \frac{w\pi(\delta_L)\sqrt{\lambda_L}}{\delta_L} + c(\lambda_L + \delta_L\sqrt{\lambda_L})$$

$$(IC_H): \quad \frac{w\pi(\gamma_H)\sqrt{\lambda_H}}{\gamma_H} + c(\lambda_H + \gamma_H\sqrt{\lambda_H}) \leq \frac{w\pi(\delta_H)\sqrt{\lambda_H}}{\delta_H} + c(\lambda_H + \delta_H\sqrt{\lambda_H})$$

$$(IR_L): \quad 0 < \gamma_L$$

$$(IR_H): \quad 0 < \gamma_H$$

Second, we build the mechanism's lagrangian function $\mathcal{L}$.

$$\mathcal{L}(\gamma_L,\gamma_H,\alpha_1,\alpha_2) = q\left[\frac{w\pi(\gamma_L)\sqrt{\lambda_L}}{\gamma_L} + (c+r)(\lambda_L + \gamma_L\sqrt{\lambda_L}) \right] \tag{19}$$
$$+ (1-q)\left[\frac{w\pi(\gamma_H)\sqrt{\lambda_H}}{\gamma_H} + (c+r)(\lambda_H + \gamma_H\sqrt{\lambda_H}) \right]$$

$$+ \alpha_1 \left[\frac{w\pi(\gamma_L)\sqrt{\lambda_L}}{\gamma_L} + c(\lambda_L + \gamma_L\sqrt{\lambda_L}) - \left(\frac{w\pi(\delta_L)\sqrt{\lambda_L}}{\delta_L} + c(\lambda_L + \delta_L\sqrt{\lambda_L}) \right) \right]$$

$$+ \alpha_2 \left[\frac{w\pi(\gamma_H)\sqrt{\lambda_H}}{\gamma_H} + c(\lambda_H + \gamma_H\sqrt{\lambda_H}) - \left(\frac{w\pi(\delta_H)\sqrt{\lambda_H}}{\delta_H} + c(\lambda_H + \delta_H\sqrt{\lambda_H}) \right) \right] \Bigg]$$

Where α_1 and α_2 are the lagrange multipliers for IC_L and IC_H, respectively. Third, we derive the first order conditions (FOC) of (19).

$$\frac{\partial \mathcal{L}}{\partial \gamma_L} = (q + \alpha_1) \left[\frac{w\pi'(\gamma_L)}{\gamma_L} - \frac{w\pi(\gamma_L)}{\gamma_L^2} + c \right] \sqrt{\lambda_L} + qr\sqrt{\lambda_L} - \alpha_2 \left[\frac{w\pi'(\delta_H)}{\delta_H} - \frac{w\pi(\delta_H)}{\delta_H^2} + c \right] \sqrt{\lambda_L} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \gamma_H} = (1 - q + \alpha_2) \left[\frac{w\pi'(\gamma_H)}{\gamma_H} - \frac{w\pi(\gamma_H)}{\gamma_H^2} + c \right] \sqrt{\lambda_H} + (1-q)r\sqrt{\lambda_H} - \alpha_1 \left[\frac{w\pi'(\delta_L)}{\delta_L} - \frac{w\pi(\delta_L)}{\delta_L^2} + c \right] \sqrt{\lambda_H} = 0$$

Letting $f(\gamma) = \frac{w\pi'(\gamma)}{\gamma} - \frac{w\pi(\gamma)}{\gamma^2} + c$ and dividing by $\sqrt{\lambda_L}$ in the first equation above and dividing by $\sqrt{\lambda_H}$ in the second equation above, we can re-write the FOC as:

$$\frac{\partial \mathcal{L}}{\partial \gamma_L} = (q + \alpha_1)f(\gamma_L) - \alpha_2 f(\delta_H) + qr = 0 \tag{20}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma_H} = (1 - q + \alpha_2)f(\gamma_H) - \alpha_1 f(\delta_L) + (1-q)r = 0 \tag{21}$$

Note that

$$\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_L} = f(\gamma_L)\sqrt{\lambda_L} \tag{22}$$

$$\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_L} = (f(\gamma_L) + r)\sqrt{\lambda_L} \tag{23}$$

Replacing (22) and (23) in (20) and (21) we can re-write the FOC as:

$$q\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_L} + \alpha_1 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_L} - \alpha_2 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma = \delta_H} = 0 \tag{24}$$

$$(1 - q)\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_H} + \alpha_2 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma = \gamma_H} - \alpha_1 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma = \delta_L} = 0 \tag{25}$$

Also, note that

$$\delta_L = \frac{\lambda_H + \gamma_H\sqrt{\lambda_H} - \lambda_L}{\sqrt{\lambda_L}} \tag{26}$$

$$\delta_H = \frac{\lambda_L + \gamma_L\sqrt{\lambda_L} - \lambda_H}{\sqrt{\lambda_H}} \tag{27}$$

follow from equation (7). These equations and the fact that $\lambda_L < \lambda_H$ imply that

$$\gamma_H < \delta_L \tag{28}$$

$$\delta_H < \gamma_L \tag{29}$$

Combining the definitions of δ_L and δ_H we get the useful relation:

$$(\delta_L - \gamma_L)\sqrt{\lambda_L} = (\gamma_H - \delta_H)\sqrt{\lambda_H} \tag{30}$$

Equation (30) implies the set of conditions:

$$\gamma_L = \delta_L \text{ if and only if } \gamma_H = \delta_H \tag{31}$$

$$\gamma_L < \delta_L \text{ if and only if } \delta_H < \gamma_H$$

$$\delta_L < \gamma_L \text{ if and only if } \gamma_H < \delta_H$$

Fourth, we characterize the natural induced truth telling region stated in the proposition.

<u>Characterization of R1:</u> Suppose $0 < \alpha_1$ and $0 < \alpha_2$. Note that $0 < \alpha_1$ implies $C_{Prog}(\gamma_L) = C_{Prog}(\delta_L)$ and $0 < \alpha_2$ implies $C_{Prog}(\gamma_H) = C_{Prog}(\delta_H)$.

There are three mutually exclusive and collectively exhaustive possibilities: either 1) $\delta_L < \bar{\gamma}$, or 2) $\delta_L = \bar{\gamma}$, or 3) $\bar{\gamma} < \delta_L$.

<u>Subcase 1:</u> $\delta_L < \bar{\gamma}$.

Note that $\delta_L < \bar{\gamma}$ implies that either $\delta_L < \bar{\gamma} < \gamma_L$ or $\delta_L = \gamma_L$. First, suppose that $\delta_L < \bar{\gamma} < \gamma_L$. Then, $\gamma_H < \delta_H$ (due to condition (31)). It also implies $\gamma_H < \delta_L < \bar{\gamma}$, and $C_{Prog}(\delta_L) < C_{Prog}(\gamma_H)$ follows from the fact that the Program's cost function is well behaved.

For $C_{Prog}(\gamma_H) = C_{Prog}(\lambda_H)$ to be true it must be that $\bar{\gamma} < \delta_H$. Also, $\delta_H < \gamma_L$ (from condition (28)), which implies that $C_{Prog}(\delta_H) < C_{Prog}(\gamma_L)$. But $C_{Prog}(\gamma_L) = C_{Prog}(\delta_L)$ (because we supposed $0 < \alpha_1$). It follows that $C_{Prog}(\delta_H) < C_{Prog}(\gamma_L) = C_{Prog}(\delta_L) < C_{Prog}(\gamma_H)$, which is a contradiction since $\alpha_2 > 0$ implies $C_{Prog}(\delta_H) = C_{Prog}(\gamma_H)$.

Second, suppose that $\delta_L = \gamma_L$. This implies $\gamma_H = \delta_H$ (from condition (31)), and $\gamma_H < \gamma_L$ follows (from condition (29)) and the fact that $\delta_L = \gamma_L$). This also implies $\delta_L = \gamma_L < \bar{\gamma}$ (from condition $0 < \alpha_1$). Using these facts in the first order conditions (24) and (25) and adding them we get:

$$q\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_L} + (1-q)\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_H} = 0 \tag{32}$$

Equation (32) only holds in three possible cases: 1) when $\gamma_L = \gamma_H = \gamma^*$, contradicting the fact $\gamma_H < \gamma_L$; 2) when $\gamma_L < \gamma^* < \gamma_H$, contradicting the fact that $\gamma_H < \gamma_L$, and 3) when $\gamma_H < \gamma^* < \gamma_L$. This case is feasible and leads to the condition

$$\delta_H = \gamma_H < \gamma^* < \gamma_L = \delta_L < \bar{\gamma} \tag{33}$$

<u>Subcase 2:</u> $\delta_L = \bar{\gamma}$.

This implies $\gamma_L = \delta_L = \bar{\gamma}$ (since the Program cost function is unimodal). It also implies that $\gamma_H = \delta_H < \bar{\gamma}$. Using these facts in the first order conditions (24) and (25) and adding them we get again equation (32). As in the previous subcase, this equation only holds in three possible cases: 1) when $\gamma_L = \gamma_H = \gamma^*$, contradicting the fact $\gamma_H < \gamma_L$; 2) when $\gamma_L < \gamma^* < \gamma_H$, contradicting the fact that $\gamma_H < \gamma_L$; and 3) when $\gamma_H < \gamma^* < \gamma_L$. This leads to the condition:

$$\delta_H = \gamma_H < \gamma^* < \gamma_L = \delta_L = \bar{\gamma} \tag{34}$$

<u>Subcase 3:</u> $\bar{\gamma} < \delta_L$. The proof for this case follows the same logic that the one for subcase 1. Note that $\bar{\gamma} < \delta_L$ implies either that $\gamma_L < \bar{\gamma} < \delta_L$ or $\bar{\gamma} < \delta_L = \gamma_L$. First, suppose that $\gamma_L < \bar{\gamma} < \delta_L$. This implies $\delta_H < \gamma_H$ (from condition 31) and $\delta_H < \gamma_L < \bar{\gamma}$. It follows that $C_{Prog}(\gamma_L) < C_{Prog}(\delta_H)$ (given that the Program cost function is well behaved). Since the cost function of the program is unimodal, for $C_{Prog}(\delta_H) = C_{Prog}(\gamma_H)$ to hold it must be that $\bar{\gamma} < \gamma_H < \delta_L$. This implies $C_{Prog}(\gamma_H) < C_{Prog}(\delta_L)$. Hence we have $C_{Prog}(\gamma_H) < C_{Prog}(\delta_L) = C_{Prog}(\gamma_L) < C_{Prog}(\delta_H)$, which contradicts the condition $C_{Prog}(\gamma_H) = C_{Prog}(\delta_H)$

(because we supposed $0 < \alpha_2$). Second, suppose that $\bar{\gamma} < \delta_L = \gamma_L$. This implies $\delta_H = \gamma_H$. By using those relations in first order conditions (24) and (25) and adding them, we get equation (32). Following the same reasoning that we used in the two previous subcases, we find the following condition:

$$\delta_H = \gamma_H < \gamma^* < \bar{\gamma} < \gamma_L = \delta_L \qquad (35)$$

We can combine conditions (33), (34) and (35) in the following condition, **which characterizes R1 in Proposition 4**:

$$\delta_H = \gamma_H < \gamma^* < \gamma_L = \delta_L \qquad (36)$$

Note that condition (36) implies that $F_L = F_H$ in region 1, which follows from replacing $\delta_L = \gamma_L$ in equation (26). Depending on parameter values, to make the low type Program indifferent between telling the truth and lying the Headquarter has to sacrifice some cost for the high type Program.

Characterization of Regions 2 and 3: Suppose that $0 < \alpha_1$ and $\alpha_2 = 0$. FOC (24) and (25) become:

$$q\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_L} + \alpha_1 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_L} = 0 \qquad (37)$$

$$(1-q)\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_H} - \alpha_1 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma=\delta_L} = 0 \qquad (38)$$

Condition (37) holds when $\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_L}$ and $\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_L}$ have opposite signs. This only happens when:

$$\gamma^* < \gamma_L < \bar{\gamma} \qquad (39)$$

Condition (38) holds when both $\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_H}$ and $\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\Big|_{\gamma=\delta_L}$ have the same sign. This happens in two mutually exclusive cases: 1) $\gamma_H < \gamma^*$ and $\delta_L < \bar{\gamma}$ simultaneously or, 2) $\gamma^* < \gamma_H$ and $\bar{\gamma} < \delta_L$ simultaneously.

First, combining condition (39) with $\gamma_H < \gamma^*$ and $\delta_L < \bar{\gamma}$ simultaneously we get $\gamma^* < \gamma_L < \bar{\gamma}$, $\gamma_H < \gamma^*$ and $\delta_L < \bar{\gamma}$. We supposed $0 < \alpha_1$. This implies $C_{Prog}(\gamma_L) = C_{Prog}(\delta_L)$. We also supposed $\alpha_2 = 0$ which implies $C_{Prog}(\gamma_H) < C_{Prog}(\delta_H)$. Since $\gamma_L < \bar{\gamma}$ and $\delta_L < \bar{\gamma}$, it must be that $\gamma_L = \delta_L$. This implies $\gamma_H = \delta_H$ (from one of the conditions 31) and $C_{Prog}(\gamma_H) = C_{Prog}(\delta_H)$ follows, contradicting $C_{Prog}(\gamma_H) < C_{Prog}(\delta_H)$.

Second, combining condition (39) with $\gamma^* < \gamma_H$ and $\bar{\gamma} < \delta_L$ simultaneously we get $\gamma^* < \gamma_L < \bar{\gamma}$, $\gamma^* < \gamma_H$, and $\bar{\gamma} < \delta_L$. Remember the definition of $\tilde{\gamma}_L$ presented in (15). It must be that $\bar{\gamma} < \tilde{\gamma}_L = \delta_L$. Otherwise, the low type Program would claim high transportation needs violating the revelation principle. It follows that: $\gamma^* < \gamma_L < \bar{\gamma} < \tilde{\gamma}_L = \delta_L$, $\gamma^* < \gamma_H$ and $\delta_H < \gamma_H$. These conditions can be divided in three sub-cases. The first Sub-case 1 is $\gamma^* < \gamma_H < \gamma_L < \bar{\gamma} < \tilde{\gamma}_L = \delta_L$, $\delta_H < \gamma_H$. For low values of q, these two conditions **characterize R2 in Proposition 4.**

The second sub-case is $\gamma^* < \gamma_L < \gamma_H < \bar{\gamma} < \tilde{\gamma}_L = \delta_L$, $\delta_H < \gamma_L$. The third sub-case is $\gamma^* < \gamma_L < \bar{\gamma} < \gamma_H < \tilde{\gamma}_L = \delta_L$, $\delta_H < \gamma_L$. For high values of q **sub-cases two and three characterize R2 in Proposition 4**.

Next, we show the existence of the threshold T_1 in Proposition 4. Let $g(\gamma) = \frac{w\pi'(\gamma)}{\gamma} - \frac{w\pi(\gamma)}{\gamma^2} + (c+r)$. Note that:

$$qg(\gamma_L)\sqrt{\lambda_L} + (1-q)g(\gamma_H)\sqrt{\lambda_H}\left(\frac{\partial \gamma_H}{\partial \gamma_L}\right) = 0 \qquad (40)$$

is a required condition for the FOC of the mechanism. The justification is as follows.

First, for *R1* we know that $\gamma_L = \delta_L$ and $\gamma_H = \delta_H$. Replacing these values in FOC (20) and (21) we get:

$$(q + \alpha_1)f(\gamma_L) - \alpha_2 f(\gamma_H) + qr = 0$$

$$(1 - q + \alpha_2)f(\gamma_H) - \alpha_1 f(\gamma_L) + (1 - q)r = 0$$

Solving the system for α_1 and α_2 we get:

$$q(f(\gamma_L) + r) + (1 - q)(f(\gamma_H) + r) = 0 \tag{41}$$

The equivalence between (40) and (41) follows from the fact that in (41) $\frac{\partial \gamma_H}{\partial \gamma_L} = \frac{\sqrt{\lambda_L}}{\sqrt{\lambda_H}}$. Second, for *R2* we know that $\alpha_2 = 0$. Hence, FOC (20) and (21) become:

$$\frac{\partial \mathcal{L}}{\partial \gamma_L} = (q + \alpha_1)f(\gamma_L) + qr = 0$$

$$\frac{\partial \mathcal{L}}{\partial \gamma_H} = (1 - q)f(\gamma_H) - \alpha_1 f(\delta_L) + (1 - q)r = 0$$

Then we get $\alpha_1 = -\frac{q(f(\gamma_L) + r)}{f(\gamma_L)}$ such that:

$$(1 - q)f(\gamma_H) + \frac{q(f(\gamma_L) + r)}{f(\gamma_L)}f(\delta_L) + (1 - q)r = 0$$

$$q(f(\gamma_L) + r) + (1 - q)(f(\gamma_H) + r)\frac{f(\gamma_L)}{f(\delta_L)} = 0 \tag{42}$$

Remember the relation $\delta_L = \tilde{\gamma}_L$ in *R2*. Using this we get $\gamma_H = \frac{\lambda_L + \tilde{\gamma}_L \sqrt{\lambda_L} - \lambda_H}{\sqrt{\lambda_H}}$. We can write $\frac{\partial \gamma_H}{\partial \gamma_L} = \frac{\partial \gamma_H}{\partial \tilde{\gamma}_L}\frac{\partial \tilde{\gamma}_L}{\partial \gamma_L}$. Note that $\frac{\partial \gamma_H}{\partial \tilde{\gamma}_L} = \frac{\sqrt{\lambda_L}}{\sqrt{\lambda_H}}$ and $\frac{\partial \tilde{\gamma}_L}{\partial \gamma_L} = \frac{f(\gamma_L)}{f(\tilde{\gamma}_L)}$. The equivalence between (40) and (42) follows.

Keeping λ_L fixed, the next step is obtaining $\frac{dC}{d\lambda_H}$.

$$\frac{dC}{d\lambda_H} = (1 - q)\left[\frac{w\pi(\gamma_H)}{2\gamma_H\sqrt{\lambda_H}} + (c + r)\left(1 + \frac{\gamma_H}{2\sqrt{\gamma_H}}\right) + g(\gamma_H)\sqrt{\lambda_H}\left(\frac{\partial \gamma_H}{\partial \lambda_H}\right)\right]$$

$$+ \frac{\partial \gamma_L}{\partial \lambda_H}\left[qg(\gamma_L)\sqrt{\lambda_L} + (1 - q)g(\gamma_H)\sqrt{\lambda_H}\left(\frac{\partial \gamma_H}{\partial \gamma_L}\right)\right]$$

Note that from condition (40) the second part of $\frac{dC}{d\lambda_H}$ equals zero. Hence,

$$\frac{dC}{d\lambda_H} = (1 - q)\left[\frac{w\pi(\gamma_H)}{2\gamma_H\sqrt{\lambda_H}} + (c + r)\left(1 + \frac{\gamma_H}{2\sqrt{\lambda_H}}\right) + g(\gamma_H)\sqrt{\lambda_H}\left(\frac{\partial \gamma_H}{\partial \lambda_H}\right)\right] \tag{43}$$

For *R1* let $\gamma_H^1 = \frac{\lambda_L + \gamma_L^1 \sqrt{\lambda_L} - \lambda_H}{\sqrt{\lambda_H}}$. Then

$$\frac{\partial \gamma_H^1}{\partial \lambda_H} = -\frac{\lambda_L + \gamma_L\sqrt{\lambda_L}}{2(\lambda_H)^{3/2}} - \frac{1}{2\sqrt{\lambda_H}}$$

$$= -\frac{\lambda_L + \gamma_L^1\sqrt{\lambda_L} + \lambda_H}{2\lambda_H\sqrt{\lambda_H}}$$

Such that

$$g(\gamma_H)\sqrt{\lambda_H}\left(\frac{\partial \gamma_H}{\partial \lambda_H}\right) = \left(\frac{w\pi'(\gamma_H)}{\gamma_H} - \frac{w\pi(\gamma_H)}{\gamma_H^2} + c + r\right)(-\sqrt{\lambda_H})\left(\frac{\lambda_L\gamma_L^1\sqrt{\lambda_L} + \lambda_H}{2\lambda_H\sqrt{\lambda_H}}\right)$$

$$= \frac{\lambda_L + \gamma_L^1\sqrt{\lambda_L} + \lambda_H}{2\lambda_H}\left(\frac{w\pi(\gamma_H)}{\gamma_H^2} - \frac{w\pi'(\gamma_H)}{\gamma_H} - (c + r)\right)$$

Replacing for $\frac{dC}{d\lambda_H}$ and simplifying we get:

$$\frac{dC}{d\lambda_H} = (1 - q)\left[\frac{w}{\sqrt{\lambda_H}}\left(\frac{\pi'(\gamma_H^1)}{\gamma_H^1} - \frac{\pi'(\gamma_H^1)}{2}\right) + \left(\frac{w\pi(\gamma_H^1)}{(\gamma_H^1)^2} - \frac{w\pi'(\gamma_H^1)}{\gamma_H^1}\right)\right] \tag{44}$$

On the other hand, for *R2* we have $\frac{\partial \gamma_H^2}{\partial \lambda_H} = -\frac{\lambda_L + \tilde{\gamma}_L^2 \sqrt{\lambda_L}}{2(\lambda_H)^{3/2}} - \frac{1}{2\sqrt{\lambda_H}}$.

Following the same reasoning we used for *R1* and replacing γ_H^1 with γ_H^2 for *R2* we get:

$$\frac{dC}{d\lambda_H} = (1-q) \left[\frac{w}{\sqrt{\lambda_H}} \left(\frac{\pi'(\gamma_H^2)}{\gamma_H^2} - \frac{\pi'(\gamma_H^2)}{2} \right) + \left(\frac{w\pi(\gamma_H^2)}{(\gamma_H^2)^2} - \frac{w\pi'(\gamma_H^2)}{\gamma_H^2} \right) \right] \tag{45}$$

Note that $\gamma_H^1 < \gamma^* < \gamma_H^2$. If we show that $\frac{dC}{d\lambda_H}$ is decreasing in γ_H^1 this is equivalent to show that $\frac{d}{d\lambda_H}(C_1 - C_2) \geq 0$. The second part of $\frac{dC}{d\lambda_H}$ in (44) is decreasing in γ_H^1 because it is the slope of a convex function. Next we show that $\frac{\pi(\gamma)}{\gamma} - \frac{\pi'(\gamma)}{2}$ is decreasing in γ. Note that $\pi'(\gamma) = \frac{\pi(\gamma)^2}{\gamma} - \gamma\pi(\gamma) - \frac{\pi(\gamma)}{\gamma}$. Replacing we get $\frac{\pi(\gamma)}{\gamma} - \frac{\pi'(\gamma)}{2} = \frac{1}{2}\left[\frac{3\pi(\gamma)}{\gamma} + \gamma\pi(\gamma) - \frac{\pi(\gamma)^2}{\gamma} \right]$. Taking the derivative with respect to γ and simplifying we get:

$$-\left(\gamma^2 + \frac{6}{\gamma^2} + 3 \right)\pi(\gamma) + \left(\frac{6}{\gamma^2} + 3 \right)[\pi(\gamma)]^2 - \frac{2}{\gamma^2}[\pi(\gamma)]^3 \leq 0$$

So $\frac{d}{d\lambda_H}(C_1 - C_2) \geq 0$ is monotonous. Now while at $\lambda_H = \lambda_L$ we know that $C_1 < C_2$, at $\lambda_H = \lambda_L + \bar{\gamma}\sqrt{\lambda_L}$ we have $C_1 > C_2$. This implies that there exists a threshold T_1 such that

$$C_1 < C_2 \text{ for } \lambda_H < T_1$$

$$C_1 \geq C_2 \text{ for } \lambda_H \geq T_1 \quad \square$$

Proof of Corollary 1: **Characterization of R3**. Suppose $\alpha_1 = 0$ and $\alpha_2 = 0$. From condition (24) we get: $q\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_L} = 0$. Due to the fact that $C_{Head}(\gamma)$ is unimodal we conclude that $\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_L} = 0$ only holds for $\gamma_L = \gamma^*$, the centralized solution found in equation (5). A similar argument for condition (25) leads to $(1-q)\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_H} = 0$, which only holds for $\gamma_H = \gamma^*$. Then the solution is $\gamma_L = \gamma_H = \gamma^*$. The Headquarter proposes $\gamma_L = \gamma_H = \gamma^*$ when $\tilde{\gamma}_L < \delta_L$. The explanation is as follows. Since $C_{Prog}(\gamma)$ is unimodal with minimum in $\bar{\gamma}$, and $\gamma^* < \bar{\gamma}$ it must be that $\bar{\gamma} < \tilde{\gamma}_L$. Hence $0 < \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \tilde{\gamma}}$. Therefore, $\tilde{\gamma} < \delta_L$ implies that $C_{Prog}(\gamma^*) = C_{Prog}(\tilde{\gamma}) < C_{Prog}(\delta_L)$. On the other hand, the high type Program would report its true needs since $\delta_H < \gamma_L = \gamma^*$ and $\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma^*} < 0$ implying $C_{Prog}(\gamma^*) < C_{Prog}(\delta_H)$. $\quad \square$

Proof of Proposition 5: Suppose $\alpha_1 = 0$ and $0 < \alpha_2$. In this case the first order conditions (24) and (25) become:

$$q\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_L} - \alpha_2 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \delta_H} = 0 \tag{46}$$

$$(1-q)\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_H} + \alpha_2 \frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_H} = 0 \tag{47}$$

Condition (46) occurs when both $\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_L}$ and $\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \delta_H}$ have the same sign. This occurs when 1) $\gamma_L < \gamma^*$ and $\delta_H < \bar{\gamma}$ simultaneously, or when 2) $\gamma^* < \gamma_L$ and $\bar{\gamma} < \delta_H$ simultaneously.

On the other hand, condition (47) occurs when $\frac{\partial C_{Head}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_H}$ and $\frac{\partial C_{Prog}(\gamma)}{\partial \gamma}\big|_{\gamma = \gamma_H}$ have opposite signs. This only occurs when $\gamma^* < \gamma_H < \bar{\gamma}$.

Combining the first case derived from condition (46) with the case derived from condition (47) we get $\gamma_L < \gamma^*$, $\delta_H < \bar{\gamma}$, and $\gamma^* < \gamma_H < \bar{\gamma}$. From condition (29) we know that $\delta_H < \gamma_L$. This implies that $\delta_H < \gamma_L < \gamma^* < \gamma_H < \bar{\gamma}$. Since $\delta_H < \bar{\gamma}$, $\gamma_H < \bar{\gamma}$ and $C_{Prog}(\delta_H) = C_{Prog}(\gamma_H)$ it must be that $\delta_H = \gamma_H$, contradicting $\delta_H < \gamma_L < \gamma^* < \gamma_H < \bar{\gamma}$.

Combining the second case derived from condition (46) with the case derived from condition (47) we get $\gamma^* < \gamma_L$, $\bar{\gamma} < \delta_H$, and $\gamma^* < \gamma_H < \bar{\gamma}$. This can be re-written as $\gamma^* < \gamma_H < \bar{\gamma} < \delta_H < \gamma_L$, which implies $\delta_L < \tilde{\gamma}_L < \bar{\gamma} < \gamma_L$.

Note that $0 < \alpha_2$ implies $C_{Prog}(\delta_H) = C_{Prog}(\gamma_H)$. Also, $\bar{\gamma} < \delta_H < \gamma_L$ implies $C_{Prog}(\delta_H) < C_{Prog}(\gamma_L)$. On the other hand, $\alpha_1 = 0$ implies $C_{Prog}(\gamma_L) = C_{Prog}(\tilde{\gamma}_L) < C_{Prog}(\delta_L)$. This implies $\delta_L < \tilde{\gamma} < \bar{\gamma} < \gamma_L$. But $C_{Prog}(\delta_H) < C_{Prog}(\gamma_L)$. This implies $\tilde{\gamma}_L < \gamma_H < \bar{\gamma}$. Then we have $\delta_L < \tilde{\gamma}_L < \gamma_H < \bar{\gamma}_L$, contradicting condition (28), which states $\gamma_H < \delta_L$. $\square$

Proof of Corollary 2: In the current system without monitoring proposition 2 guarantees that the high type Program always reveals its true transportation needs. On the other hand, threshold (3) defines the truth telling condition for the low type Program. If $\tilde{\gamma}_L < \delta_L$, then the low type Program reveals its true transportation needs and gets $F(\gamma^*, \lambda_L) < F(\gamma^*, \lambda_H)$. This corresponds to region 4 in the mechanism. If $\tilde{\gamma}_L = \delta_L$, the low type Program is indifferent between revealing the truth and inflating its needs and gets $F(\delta_L, \lambda_L) = F(\gamma^*, \lambda_H)$. The same outcome can be obtained in the mechanism in region 1. Otherwise, if $\delta_L < \tilde{\gamma}_L$, then the low type Program inflates its transportation needs and gets $F(\delta_L, \lambda_L) = F(\gamma^*, \lambda_H)$. We can always replicate this outcome with the mechanism by choosing $\lambda_H < \gamma^* < \gamma_L$ such that $F(\gamma_L, \lambda_L) = F(\gamma_H, \lambda_H)$ and $F(\gamma_H, \lambda_H) < F(\gamma^*, \lambda_H)$. $\square$

References

Balcik, B., B. M. Beamon, K. Smilowitz. 2008. Last Mile Distribution in Humanitarian Relief. *Journal of Intelligent Transportation Systems* 12(2) 51–63.

Barbarosoglu, G., L. Ozdamar, A. Cevik. 2002. An interactive approach for hierarchical analysis of helicpter logistics in disaster relief operations. *European Journal of Operational Research*, 140 (1), 118–133.

Barbarosoglu, G., Y. Arda. 2004. A two-stage stochastic Programming framework for transportation planning in disaster response. *Journal of the Operational Research Society.* 55 (1), 43–53.

Batta, R., N.R. Mannur. 1990. Covering-location models for emergency situations that require multiple response units. *Management Sci.* 36 (1), 16–23.

Beamon, B.M., S.A. Kotleba. 2006. Inventory modelling for complex emergencies in humanitarian relief operations. *International Journal of Logistics: Research and Applications.* 9 (1), 1–18.

Bernstein, F., A. Federgruen. 2003. Pricing and replenishement strategies in a distribution system with competitive retailers. *Operations Research.* 51(3) 409–426.

Bookbinder, J.H., D.L. Martell. 1979. Time-dependent queueing approach to helicopter allocation for forest fire initial-attack *INFOR* 17(1) 58–70.

Borst, S. Mandelbaum, A. Reiman. 2004. Dimensioning large call centers. *Operations Research.* 52(1) 17–34.

Campbell A.M., D. Vanderbussche, W. Hermann. 2008. Routing for relief efforts. *Transportation Sci.* 42(2) 127–145.

Cachon, G.P., M.A. Lariviere. 2005. Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management Sci.* 51(1) 30–44.

Cachon, G.P., Tearwiesch, C. 2009. Matching supply with demand. An introduction to operations management. *McGraw-Hill.* Second Edition - International Edition.

Chang, M-S, Y.-L. Tseng, J-W Cheng. 2007. A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. *Transportation Research Part E.* 43 737-754

Chiu, Y.C., H. Zheng. 2007. Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: Model formulation and solution *Transportation Research Part E.* 43 710-736

Corbett, C.J., X. de Groote. 2000. A supplier's optimal quantity discount policy under asymmetric information. *Management Sci.* 46(3) 444–450.

Cova, T.J., J.P. Johnson. 2003. A network flow model for lane-based evacuation routing *Transportation Research Part A.* 37 579-604

De Angelis, V. M. Mecoli, C. Nikoi, G. Storchi. 2007. Multiperiod integrated routing and scheduling of World Food Programme cargo planes in Angola. *Computers & Operations Research.* 34 1601–1615

Disparte, D. 2007. The Postman Parallel. *CarNation.* 2 22–27.

Grassmann, W.K. 1988. Finding the right number of servers in real-world queuing systems. *Interfaces.* 18(2) 94–104.

Green, L.V.. 1984. A multiple dispatch queueing model of police patrol operations. *Management Sci.* 30(6) 653–664.

Green, L.V., P. Kolesar. 1984a. The feasibility of one-officer patrol in New York City. *Management Sci.* 30(8) 964–981.

Green, L.V., P. Kolesar. 1984b. A comparison of the multiple dispatch and M/M/c priority queueing models of police patrol. *Management Sci.* 30(6) 665–670.

Green, L.V., P. Kolesar. 1989. Testing the validity of a queueing model of a police patrol. *Management Sci.* 35(2) 127–148.

Green, L.V., P. Kolesar. 2004. Improving emergency responsiveness with management science. *Management Sci.* 50(8) 1001–1014.

Halfin, S. W. Whitt. 1981. Heavy-traffic limits of queues with many exponential servers. *Operations Research.* 29(3) 567–588

Hasija, S., E.J. Pinker, R.A. Shumsky. 2005. Stffinf and routing in a two-tier call centre. *Int. J. Operational Research.* 1(1) 8–29 .

Hasija, S., E.J. Pinker, R.A. Shumsky. 2008. Call center outsourcing contracts under information asymmetry. *Management Sci.* 54(4) 793–807.

Ignall, E., Carter, G., Rider, K. 1982. An algorithm for the unitial dispatch of fire companies. *Management Sci.* 28(4) 366-378

Jia, H., F. Ordonez, M. Dessouky. 2007. A modeling framework for facility location of medical services for large-scale emergencies *IIE Transactions.* 39 41-55

Kolesar, P., E. W. Blum. 1973. Square root laws for fire engine response distances. *Management Sci.* 19(2) 1368–1378.

Krishnan, H., R. Kapuscinski, D.A. Butz. 2007. Coordinating contracts for decentralized supply chains with retailer promotional effort. *Management Sci.* 50(1) 48–63.

Lindenberg, M. 2001. Are we at the cutting edge of the blunt blunt edge? Improving NGO organizatinal performance with private and public sector strategic management frameworks *Nonprofit Management & Leadership.* 11(3) 247–270.

Mannell, J. 2010. Are the sectors compatible? International development work and lessons for a business–profit partnership framework *Journal of Applied Social Psychology.* 40(5) 1106–1122.

Ozdamar, L., E. Ekinci, B. Kucukyazici. 2004. Emergency logistics planning in natural disasters *Annals of Operations Research.* 129 217–245

Pedraza Martinez, A.J., O. Stapleton, L.N. Van Wassenhove. 2010. Field vehicle fleet management in humanitarian operations: a cse-based approach. *INSEAD Working Paper 2010/38/TOM/ISIC.*

Pedraza Martinez, A.J., L.N. Van Wassenhove. 2010. Vehicle replacement in the International Committee of the Red Cross. *Production and Operations Management* (Forthcoming).

Pasternack, B.A. 1985. Optimal pricing and return policies for perishable commodities. *Marketing Sci.* 4(2) 166–176.

Regnier, E. 2008. Public evacuation decisions and hurricane track uncertainty. *Management Sci.* 54 (1) 16–28

Saadatseresht, M., A. Mansourian, M. Taleai. 2009. Evacuation planning using multiobjective evolutionary optimization approach. *European Journal of Operational Research.* 198 305–314

Salmeron, J., A. Apte. 2009. Stochastic optimization for natural disaster asset prepositioning. *Production and Operations Management.* 19(5) 561–574

Sheu, J.B. 2007. An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research E.* 43 687 – 709

Singer, M., P. Donoso. 2008. Assessing an ambulance service with queuing theory. *Computers & Operations Research.* 35 2549–2560

Stepanov. A., J.M. Smith. 2009. Multi-objective evacuation routing in transportation networks. *European Journal of Operational Research.* 198 435–446

Su, X., S.A. Zenios. 2006. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: a mechanism design model. *Management Sci.* 52(11) 1647–1660.

Taylor, T. 2002. Supply chain coordination under channel rebates with sales effort effects. *Management Sci.* 48(8) 992–1007.

Thomas, A., L. Kopczak. 2005. From logistics to supply chain management: the path forward in the humanitarian sector. *Fritz Institute White Paper*

Van Wassenhove, L. N. 2006. Humanitarian aid logistics: supply chain management in high gear. *Journal of the Operational Research Society* 57 475–489.

Whitt, W. 1992. Understanding the efficiency of multi-server service systems. *Management Sci.* 38(5) 708–723.

Yi, W., L. Ozdamar. 2007. A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research.* 179(3) 1177–1193.

INSEAD

The Business School
for the World®