Ann Oper Res (2010) 181: 559–589
DOI 10.1007/s10479-010-0787-x

Alternate risk measures for emergency medical service system design

Nilay Noyan

Published online: 1 September 2010

Abstract The stochastic nature of emergency service requests and the unavailability of emergency vehicles when requested to serve demands are critical issues in constructing valid models representing real life emergency medical service (EMS) systems. We consider an EMS system design problem with stochastic demand and locate the emergency response facilities and vehicles in order to ensure target levels of coverage, which are quantified using risk measures on random unmet demand. The target service levels for each demand site and also for the entire service area are specified. In order to increase the possibility of representing a wider range of risk preferences we develop two types of stochastic optimization models involving alternate risk measures. The first type of the model includes integrated chance constraints (ICCs), whereas the second type incorporates ICCs and a stochastic dominance constraint. We develop solution methods for the proposed single-stage stochastic optimization problems and present extensive numerical results demonstrating their computational effectiveness.

Keywords Stochastic programming · Random demand · Risk constraints · Integrated chance constraints · Stochastic dominance · Emergency system · Facility location · Ambulance allocation · Equity

1 Introduction

Determining the optimal location of emergency vehicles is a significant problem in designing EMS systems and has received considerable attention in the literature (Brotcorne et al. 2003; Marianov and ReVelle 1995). A key point in effective emergency response is the prompt availability of emergency vehicles at response facilities. The service area of an EMS system is often modeled by defining a network consisting of a set of geographical nodes,

N. Noyan (✉)
Manufacturing Systems/Industrial Engineering Program, Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı, Tuzla, 34956 Istanbul, Turkey
e-mail: nnoyan@sabanciuniv.edu

where each node represents a source of requests for response (emergency) vehicles it may also denote a candidate facility site.

We consider an EMS system design problem of locating the response facilities (ambulance stations) and determining the number of vehicles (ambulances) to allocate to each facility. In real life situations, the future emergency service requests from demand sites are not known with certainty. Furthermore, due to the various constraints, such as budget or capacity, pre-allocated emergency vehicles may not be sufficient to cover all demand in the service area within an acceptable time. Since unmet demands in emergency situations may result in loss of life, it is critical to design systems that guarantee reasonable levels of coverage for potential users. Hence, the demand uncertainty must be addressed effectively in designing EMS systems.

Stochastic programming is one of the fundamental approaches that can be used to model decision problems in the presence of uncertainty. Here, we model the uncertainty in demand for the emergency vehicles. In particular, we represent the uncertain demands by random variables and use a scenario approach to characterize the randomness and model risk, which can be broadly defined as the effect of variability of random demand. We describe two types of new stochastic programming formulations that determine the optimal location and allocation decisions minimizing the total cost while meeting the target service levels. The level of service is measured by keeping the unmet demand values below some prescribed target values, guaranteeing a high level of coverage. We specify an individual target service level for each demand node and a target service level for the entire geographical area (system-wide coverage). Defining only a system-wide target service level may result in a lower level of coverage at one demand site and a higher level of coverage at another site, hence leading to inequitable solutions. Providing equal access to users is an important issue in EMS system design (Felder and Brinkmann 2002). However, there is no agrement on what is fair in an EMS system and how to measure the equity. For instance, Felder and Brinkmann (2002) discuss the conflicts between the equal access policy guaranteeing a nationwide uniform response time and the policy aiming to provide equal per-capita resources in EMS systems. Felder and Brinkmann (2002) suggest that every person should not be given equal access to emergency service irrespective of locations. For further discussion on equity in locating facilities, we refer to Marsh and Schilling (1994). In this study, the way we define the individual target service levels may be regarded as an alternative approach to model the coverage equity. We consider an equal access policy which guarantees a system-wide uniform response time and also guarantee equal service at each demand site based on the proportion of the unmet demand. This implies that our models consider the level of demand as a criterion and allocate more vehicles to higher-demand areas than lower-demand areas, but in terms of proportionality different areas would get equal service. To the best of our knowledge, there does not exist an EMS design model like ours considering simultaneously individual target service levels and a system-wide service level.

Stochastic programming formulations using probabilistic constraints, also called chance constraints, are widely applied in stochastic EMS design models. These models usually consider the probability of having an available vehicle within a standard acceptable distance as the performance measure. However, Erkut et al. (2008b) argue in detail that this probabilistic performance measure is not consistent with the performance measures used by most EMS operators in practice. A common performance measure is the fraction of calls covered whose response time is below a specified threshold. Discussions by Erkut et al. (2008b) indicate the significance of the models based on the expected coverage performance measures. Following this line of thought, we propose to use alternate risk measures based on the expected unmet demand; and therefore, our proposed risk measures may potentially be better aligned with the performance measures employed by EMS operators in practice. The major contribution of our study is the use of integrated chance constraints and stochastic dominance constraints as alternatives to probabilistic constraints in EMS system design problems.

Probabilistic constraints are commonly used; however, it is well known that they pose great computational difficulties. Therefore, in general we can only solve small to moderate size problems involving probabilistic constraints. In this study, we show that switching to alternate risk constraints and developing corresponding solution methods we obtain computationally tractable models for a larger set of scenarios compared to the existing literature. Handling a larger set of scenarios is significant in modeling uncertainties of real life. In our first type of stochastic optimization model, the target service levels are defined using integrated chance constraints. Note that probabilistic constraints measure the probabilities of violating the coverage constraints, irrespective of how violated the constraints are. In other words, the probabilistic constraints do not take the magnitude of the unmet demand into account. As an alternative, ICCs are based on the magnitude of violation in coverage constraints. ICCs were introduced by Klein Haneveld (1986) and have only been used in finance applications so far. The use of ICCs in EMS design is novel. In our second type of model, the individual target service level for each demand node is defined using an ICC on the random unmet demand, whereas the system-wide service level is defined using a second order stochastic dominance (SSD) constraint introduced by Dentcheva and Ruszczyński (2003). The stochastic dominance relation allows us to obtain location and allocation decisions for which the random total unmet demand dominates a benchmark (reference) random total unmet demand. Such a reference outcome may be defined based on an EMS standard or a potential/candidate solution. In either case, there is a reference random outcome and the model involving stochastic dominance constraints constructs a decision vector for which the associated random outcome is better than the reference with respect to the specified dominance criterion. We propose to define the reference outcome based on a common EMS performance standard which imposes a lower bound on the fraction of calls whose response time is below a threshold. By proposing two models we increase the possibility of representing a wider range of risk preferences. In many applications, where the distribution of a random outcome is of significant interest and it is possible to define a reasonable reference random outcome, we recommend the decision makers to use the model involving stochastic dominance constraints.

The literature review is presented in Sect. 2. In Sect. 3, we describe the underlying deterministic problem for the proposed stochastic programming models. In Sect. 4, we describe the general framework, which incorporates the risk constraints into the single- and two-stage stochastic optimization models. After discussing the computational challenge of the two-stage formulations, we restrict our attention to the single-stage formulations for the remainder of the paper. In Sect. 5, we first introduce a mixed integer linear programming (MILP) formulation for the single-stage stochastic EMS system design model with ICCs and then develop an associated computationally effective alternate formulation. In Sect. 6, a single-stage stochastic optimization problem involving ICCs and an SSD constraint, more precisely an “increasing convex order” constraint, is presented. We also describe effective and practical methods to solve this problem. We present numerical results in Sect. 7 to demonstrate the computational effectiveness of the developed solution methods and illustrate how input parameters and risk measures affect the optimal location and allocation decisions. In Sect. 7, we also discuss the computational study performed to compare the proposed models with a closely related existing model. Finally, in Sect. 8 we conclude and discuss further research directions.

2 Literature review

The problem of determining the optimal locations of emergency vehicles has been quite popular in the MS/OR literature. For extensive reviews on emergency vehicle location and allocation models we refer to Brotcorne et al. (2003), Marianov and ReVelle (1995), and Goldberg (2004). EMS design problems are closely related to the facility location problems. The readers should consult the book by Daskin (1995) and the references therein for a detailed discussion on facility location theory. The models we propose in this study are stochastic versions of the capacitated facility location problem. Our aim is not to provide an extensive review; but to briefly discuss some selected relevant papers on modeling the uncertainty in location problems. Comprehensive reviews on facility location under uncertainty can be found in Berman and Krass (2001), Louveaux (1993), Owen and Daskin (1998), and Snyder (2006). One of the first probabilistic models for locating emergency vehicles, known as the maximum expected covering location model (MECLM), is proposed by Daskin (1983). MECLM accounts for the potential unavailability of ambulances and maximizes the expected demand coverage for a given number of facilities to be located on the network under the assumption that each ambulance has the same probability of being unavailable to serve a call. As a generalization of the maximum covering model, ReVelle and Hogan (1989) propose chance constrained stochastic models which maximize the demand covered with a given probability value. There is also a rich literature on emergency vehicle location models focusing on randomness in response times (e.g., see Ingolfsson et al. 2008; Erkut et al. 2008a). After this brief review of stochastic location models in general, we next focus on stochastic versions of the classical capacitated fixed charge facility location problem (CFLP) which are particularly related to our study.

Research focusing on the stochastic CFLP includes Louveaux (1986), Ball and Lin (1993), Beraldi et al. (2004) and Beraldi and Bruni (2009). Louveaux (1986) presents a stochastic version of the CFLP in which the expected utility of profit is maximized while considering a penalty for unmet demand. Ball and Lin (1993), Beraldi et al. (2004) and Beraldi and Bruni (2009) assume that the main uncertainty is due to the stochastic call arrival process, and they propose stochastic programming formulations under probabilistic constraints. Prékopa (1995) discusses in detail the probabilistic optimization theory and the associated numerical techniques. Ball and Lin (1993) incorporate a probabilistic constraint for each demand site to ensure that the probability of unavailability of a vehicle to serve a request from the demand site within an acceptable time is less than a certain value. On the other hand, Beraldi et al. (2004) and Beraldi and Bruni (2009) incorporate probabilistic constraints to ensure that all requests are served with a prescribed high probability. Since Beraldi et al. (2004) and Beraldi and Bruni (2009) directly focus on the randomness in demand satisfaction rather than the randomness in the availability of vehicles, these studies are more closely related to our study than Ball and Lin (1993) and we would like to discuss our contribution relative to these studies. Beraldi and Bruni (2009) introduce a two-stage stochastic programming problem, where the second stage decision variables are associated with scenarios to represent the assignment of vehicles to demand nodes under each scenario. In this study, we do not consider scenario dependent vehicle assignment decisions and only focus on locating the response facilities and determining the number of vehicles to allocate to each facility. Thus, we mainly focus on single-stage stochastic programming models similar to Beraldi et al. (2004). Moreover, Beraldi and Bruni (2009) do not allow splittable demand, i.e., the demand at each node must be served by exactly one facility under each scenario. Therefore, our paper has more commonalities with Beraldi et al. (2004) and we performed a computational study to compare our results to those that would be obtained by using probabilistic constraints as in Beraldi et al. (2004).

Some previous studies like Daskin (1983) and Revelle and Hogan (1989) assume that the service providers operate independently. Although calls for service may arrive independently, the assumption of independence among service providers may not be justified and relaxing this independence assumption is crucial to handle the spatial dependencies of demand sites (e.g., see Batta et al. 1989). We model the random demands using the scenario approach and relax the assumptions that the service providers operate independently and the demand sites are independent. Beraldi et al. (2004) assume that the demand distributions at each node are given and random demands are independent. Under the assumption of independent demands they reformulate the problem involving probabilistic constraints and solve it using the CPLEX solver. Without this independence assumption it is hard to solve such a problem and the reformulation presented in Beraldi et al. (2004) is not valid for a given set of scenarios characterizing the joint demand. Therefore, a different reformulation is required to incorporate the probabilistic constraints when a scenario approach is used to model risk. These discussions support the potential contribution of the proposed risk constraints alternate to probabilistic constraints.

In general, one cannot claim that one risk measure is better than others. Depending on the decision maker's risk preference, either probabilistic constraint, or ICCs, or stochastic dominance constraints may be employed. However, our motivation to propose risk constraints alternate to probabilistic constraints in EMS system design models is based on the arguments by Erkut et al. (2008b). Apart from this, the computational difficulties inherent in probabilistic constraints further motivate our research. The risk measures we propose to model different types of risk preferences, lead to computationally tractable single-stage models and allow us to handle a larger set of scenarios for improving the validity of the models.

3 The underlying deterministic model

In this section we present our assumptions and notation, and describe the underlying deterministic model for the stochastic programming models that are presented in the following sections. Then, we discuss how to incorporate stochastic demand into the model.

We say that a candidate facility can cover a demand node if the distance between them is less than or equal to an acceptable value, which is known as the coverage distance threshold and can be determined according to a response time standard.

Inputs

I: finite set of demand sites (nodes);
J: finite set of candidate facility sites, where response facilities can be located;
d_{ij}: traveling distance between demand node i and candidate facility node j, $i \in I$, $j \in J$;
T_c: coverage distance threshold;
$M_j = \{i \in I \mid d_{ij} \leq T_c\}$: set of demand nodes that can be covered by a facility located at node j, $j \in J$;
$N_i = \{j \in J \mid d_{ij} \leq T_c\}$: set of candidate facility nodes within acceptable distance of node i, $i \in I$;
f_j: (hourly) fixed cost of opening a facility at node j, $j \in J$;
a: (hourly) cost of purchasing and maintaining an emergency vehicle;
β: cost of shipping per unit distance per unit demand;
c_{ij}: cost of shipping a unit demand from a facility at node j to node i (notice that $c_{ij} = \beta d_{ij}$), $i \in I$, $j \in J$;

U_j: maximum number of vehicles that can be assigned to a facility located at node j, $j \in J$;
h_i: number of service requests (demand) generated at node i, $i \in I$, during a specified amount of time.

Decision variables

$$y_j = \begin{cases} 1 & \text{if a facility is located at node } j \in J, \\ 0 & \text{otherwise.} \end{cases}$$

x_{ij} = Number of vehicles located at facility site $j \in J$ due to demand at node $i \in I$.

We remark that we do not explicitly model how an available response vehicle is assigned to an emergency call, i.e., we do not consider any particular dispatching rule. The allocation variable x_{ij} is interpreted as the number of vehicles located at facility site j due to the demand at node i. However, *these vehicles are not dedicated to node i* in real life dispatching problems. In other words, $\sum_{i \in M_j} x_{ij}$ only represents the total number of vehicles allocated to node j and these vehicles are not reserved to serve specific demand nodes. For the determined facility locations and the number of vehicles allocated to each facility, existing methods may be utilized to find practical solutions for dispatching and reallocating vehicles.

We consider service requests at each demand site during a certain amount of time. The length of this time period is chosen as a reasonable time required for a service trip, which we define as the total time required for an emergency vehicle to return to the original location before responding to another service request. Similar to other studies in this area, such as Beraldi et al. (2004) and Beraldi and Bruni (2009) we consider hourly demand.

For our stochastic programming problems, we have the following *underlying deterministic problem*:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j + \sum_{i \in I} \sum_{j \in J} a x_{ij} \tag{1}$$

$$\text{subject to:} \quad \sum_{j \in N_i} x_{ij} \geq h_i, \quad \forall i \in I, \tag{2}$$

$$\sum_{i \in M_j} x_{ij} \leq U_j y_j, \quad \forall j \in J, \tag{3}$$

$$x_{ij} \in \mathbb{Z}_+, \quad \forall i \in I, \ j \in J, \tag{4}$$

$$y_j \in \{0, 1\}, \quad \forall j \in J. \tag{5}$$

This is similar to the well known mathematical programming formulation of the capacitated fixed charge facility location problem with splittable demands. The objective function (1) minimizes the sum of the variable transportation costs, the fixed setup cost for opening the facilities and the total cost of purchasing and maintaining vehicles. Different than the traditional CFLP, we also incorporate the total ambulance cost which is obtained by multiplying the total number of ambulances by the unit cost a. In ambulance service, 24-hour availability is required and therefore, it has relatively high fixed costs associated with equipment and staffing. However, the variable costs based on distances are significant in allocating vehicles since the closest available vehicle is usually dispatched to an emergency call. Each demand node i has a total request of h_i vehicles, and coverage constraints (2) ensure that all of the demand at each node must be served by one or more open facilities. Constraints (3) are capacity constraints, which guarantee that the amount of vehicles allocated to each facility site $j \in J$ is less than or equal to the maximum number of vehicles that can be assigned to facility node j. Constraints (4) and (5) are the integrality and nonnegativity constraints. We denote the cardinality of a set A by $|A|$. Notice that $x_{ij} = 0$ for $j \notin N_i,\ i \in I$, and therefore, the dimension of the decision vector $\mathbf{x}$ reduces to $\sum_{i \in I} |N_i|$.

In practice, as we determine the values of the allocation vector $\mathbf{x}$ and the location vector $\mathbf{y}$, the actual values of $h_i,\ i \in I$, are not known; they will become known in the future. In this paper, we consider models where the demand parameters, $h_i,\ i \in I$, are not constants but random variables denoted by $h_i(\omega),\ i \in I$. This implies that coverage constraints (2) are stochastic. In the following Sects. 4, 5 and 6, we develop mathematical programming formulations, which involve risk measures to model the uncertainty in demand using the scenario approach.

4 General framework with risk constraints

We model risk, which can be broadly defined as the effect of variability of random service requests, by specifying constraints on random unmet demands. The underlying idea of the models we consider is to allow infeasibilities in the stochastic version of the coverage constraints (2), while specifying restricting constraints on the amount of their violations. We assume that we are given a discrete set of scenarios, a set of realizations of joint service requests at the demand nodes, and their associated probabilities. Let S denote the finite set of (global) scenarios, p_s denote the probability associated with scenario $s,\ s \in S$, and h_i^s denote the realization of demand at node i under scenario s, $i \in I, s \in S$. It is worthwhile to point out that using scenarios allows the demand values to be dependent.

The general formulation of the proposed single-stage stochastic EMS design optimization models reads:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j + \sum_{i \in I} \sum_{j \in J} a x_{ij} \tag{6}$$

$$\text{subject to:} \quad \text{risk constraints on the violation of } \left(\sum_{j \in N_i} x_{ij} \geq h_i(\omega), \quad \forall i \in I \right), \tag{7}$$

$$\sum_{i \in M_j} x_{ij} \leq U_j y_j, \quad \forall j \in J, \tag{8}$$

$$x_{ij} \in \mathbb{Z}_+, \quad \forall i \in I,\ j \in J, \tag{9}$$

$$y_j \in \{0, 1\}, \quad \forall j \in J. \tag{10}$$

As an alternative approach, we can define a decision variable x_{ij} associated with each scenario, denoted by $x_{ij}^s,\ s \in S$, and determine how to assign the vehicles to demand nodes under each scenario. Basically, $x_{ij}^s,\ i \in I,\ j \in J,\ s \in S$, are the second-stage decisions, which depend on the realized values of the random demands, and we can state that they are the realizations of the random decision variables $x_{ij}(\omega),\ i \in I,\ j \in J$. Then the traditional two-stage stochastic programming formulation of the EMS design problem, presented in Sect. 3, is given by

$$\min \sum_{j \in J} f_j y_j + \sum_{j \in J} a v_j + E[Q(\mathbf{y}, v, \mathbf{h}(\omega))] \tag{11}$$

$$\text{subject to:} \quad \upsilon_j \leq U_j y_j, \quad \forall j \in J, \tag{12}$$

$$y_j \in \{0, 1\}, \quad \forall j \in J, \tag{13}$$

$$\upsilon_j \in \mathbb{Z}_+, \quad \forall j \in J, \tag{14}$$

where υ_j denotes the number of vehicles allocated to facility node $j,\ j \in J$. For the realization of the random demand vector $\mathbf{h}(\omega)$ under scenario $s,\ s \in S$, the second-stage problem is given by

$$Q(\mathbf{y}, \upsilon, \mathbf{h}^s) = \min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij}^s \tag{15}$$

$$\text{subject to:} \quad \sum_{j \in N_i} x_{ij}^s \geq h_i^s, \quad \forall i \in I, \tag{16}$$

$$\sum_{i \in M_j} x_{ij}^s \leq \upsilon_j, \quad \forall j \in J, \tag{17}$$

$$x_{ij}^s \in \mathbb{Z}_+, \quad \forall i \in I,\ j \in J. \tag{18}$$

Unlike the traditional two-stage stochastic programming, in our proposed approach we allow constraints (16) to be violated and we impose risk constraints on the solvability of the second-stage problem. Then the two-stage version of the proposed general formulation (6)–(10) becomes

$$\min \sum_{j \in J} f_j y_j + \sum_{j \in J} a \upsilon_j + \sum_{s \in S} \sum_{i \in I, j \in J} p_s c_{ij} x_{ij}^s \tag{19}$$

$$\text{subject to:} \quad \text{risk constraints on the violation of } \left(\sum_{j \in N_i} x_{ij}(\omega) \geq h_i(\omega), \quad \forall i \in I \right), \tag{20}$$

$$\sum_{i \in M_j} x_{ij}^s \leq \upsilon_j, \quad \forall j \in J,\ s \in S, \tag{21}$$

$$\upsilon_j \leq U_j y_j, \quad \forall j \in J, \tag{22}$$

$$y_j \in \{0, 1\}, \quad \forall j \in J, \tag{23}$$

$$\upsilon_j \in \mathbb{Z}_+, \quad \forall j \in J, \tag{24}$$

$$x_{ij}^s \in \mathbb{Z}_+, \quad \forall i \in I,\ j \in J,\ s \in S. \tag{25}$$

It is easy to see that $\upsilon_j = \max_{s \in S} \sum_{i \in M_j} x_{ij}^s$ for all $j \in J$.

Introducing risk constraints on the solvability of the second-stage problem has been first proposed by Prékopa (1973). Mainly probabilistic constraints are considered in such a framework (e.g., see Prékopa 1973; Noyan and Prékopa 2006; Beraldi and Bruni 2009). In Sects. 5 and 6, we develop single-stage stochastic programming formulations, which incorporate alternate risk measures to specify the risk constraints on the random unmet demands. The risk constraints proposed for the single-stage formulations directly apply to their two-stage counterparts (20) if the variables x_{ij} are replaced by their scenario-dependent versions $x_{ij}^s,\ s \in S$. Although we do not explicitly state these two-stage formulations, we note that introducing such alternate risk measures on the solvability of the second-stage problem is a valuable contribution in its own right.

The single-stage formulations are based on the simplifying assumption that allocations do not depend on scenarios. Basically, the risk constraints in the single-stage formulations are safe approximations of those in the two-stage formulations. In other words, the two-stage formulations would find solutions with lower total costs with respect to their associated single-stage formulations under flexible allocations. However, a single-stage formulation would be the right choice in a setting where the capacity allocation needs to be specified for each demand node a priori, i.e., if x_{ij} is pre-allocated for demand node i. The ideal modeling choice clearly depends on the problem context. The downside of the two-stage formulations is the computational challenge of solving them optimally due to their size. Therefore, we choose to dedicate our study to the single-stage models because the main focus of this paper is to introduce alternate risk measures in stochastic EMS design models and deal with large problem instances. In this study, we do not focus on the assignment decisions under each scenario and define the variables x_{ij} only to represent the inter-facility resource allocation. The proposed single-stage stochastic programming models provide us with x_{ij}, $i \in I$, $j \in J$, values that are large enough to guarantee the satisfaction of demand with prescribed service levels, regardless of the realization of (actual value taken by) demand. Due to the safe approximation, policies obtained by such models are more conservative and lead to higher total costs. In Sect. 7.2, we present some illustrative numerical results to give some insights about the relative increase in the total cost due to the simplifying assumption that allocations are pre-determined.

Beraldi and Bruni (2009) consider a two-stage formulation involving a joint probabilistic constraint for a simpler version of the EMS design problem, where the demands are not splittable. According to their numerical results, it is even hard to solve their proposed two-stage stochastic formulation for instances with $|I| = 100$, $|J| = 50$, $|S| = 40$. Our proposed alternate risk measures avoid some of the computational difficulties inherent in probabilistic constraints, but it is still hard to solve the two-stage versions of the our models for a moderate number of scenarios, e.g., $|S| = 200$. In the following Sects. 5 and 6, we show that the proposed stochastic programming models can be reformulated as MILP problems. Preliminary results show that even solving the LP-relaxation of the linearized formulation of (19)–(25) is hard for large problem instances due to the number of assignment decisions ($|S| * \sum_{i\in I} |N_i|$). In this study, we are mainly interested in identifying policies for a large set of scenarios and therefore, we restrict our attention to the single-stage formulations. As a part of our ongoing research, we focus on developing methods to solve the two-stage versions of the proposed models for moderate size problems. The rest of the paper is dedicated to the single-stage stochastic programming models.

5 An integrated chance constrained EMS system design model

In Sect. 5.1, we discuss the integrated chance constraints, and then in Sect. 5.2 we introduce *the integrated chance constrained EMS system design* problem. Finally, in Sect. 5.3 we develop an equivalent alternate formulation, which leads to a significant reduction in the number of variables and provides us with an effective solution method.

5.1 Integrated chance constraints

In connection with the stochastic constraints there are several measures of violation that can be incorporated into an optimization model. One way of measuring such violations is via probabilistic constraints. Integrated chance constraints (ICCs) are introduced by Klein Haneveld (1986) as alternate to probabilistic constraints. Integrated chance constraints can be considered as relaxations of probabilistic constraints. Therefore, ICCs can be used to obtain convex approximations of the generally non-convex feasible sets defined by probabilistic constraints.

We model the risk by specifying constraints on random unmet demands and the random total unmet demand. Let $e_i : \mathbb{Z}_+{}^{|I|\times|J|} \times S \to \mathbb{R},\ i \in I$, be the outcome mappings. For a given allocation vector $\mathbf{x} \in \mathbb{Z}_+{}^{|I|\times|J|}$ let us define the mapping $e_{(\mathbf{x},i)} : S \to \mathbb{R}$ by $e_{(\mathbf{x},i)}(s) = e_i(\mathbf{x}, s)$ for all $i \in I,\ s \in S$. Also let $[\eta]_+ = \max(0, \eta)$ and $[\eta]_- = \max(0, -\eta)$ for $\eta \in \mathbb{R}$. We denote the random unmet demand at node i by $e_{(\mathbf{x},i)},\ i \in I$, and the random total unmet demand (for the network) by $\xi_{\mathbf{x}}$, where

$$e_{(\mathbf{x},i)}(s) = \left[h_i^s - \sum_{j\in N_i} x_{ij} \right]_+, \quad s \in S, \tag{26}$$

and

$$\xi_{(\mathbf{x})}(s) = \sum_{i\in I} e_{(\mathbf{x},i)}(s) = \sum_{i\in I} \left[h_i^s - \sum_{j\in N_i} x_{ij} \right]_+, \quad s \in S. \tag{27}$$

Using the definition of Klein Haneveld (1986) we have the ICCs on random unmet demands as follows:

$$\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij} \right]_+ \right) \le q_i, \quad i \in I, \tag{28}$$

where $\mathbb{E}$ stands for the expected value operator and $q_i, i \in I$, are nonnegative risk aversion parameters representing the largest acceptable expected unmet demand values. The constraints of type (28) guarantee that for all demand nodes the average magnitude of unmet demand is less than or equal to the maximum acceptable risk aversion parameters. For example, one can set $q_i = \gamma_i E[h_i(\omega)]$, where γ_i is another type of risk aversion parameter and this would mean that at most a fraction γ_i of the expected demand be unmet. Here as proposed by Klein Haneveld and Van Der Vlerk (2006), we construct alternative individual ICCs by choosing the risk parameters dependent on the distributions of random unmet demands instead of specifying the maximum acceptable risk parameters as fixed numbers $q_i, i \in I$. These alternative ICCs are

$$\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij} \right]_+ \right) \le \alpha_i \mathbb{E}\left(\left| h_i(\omega) - \sum_{j\in N_i} x_{ij} \right| \right), \quad i \in I, \tag{29}$$

where $\alpha_i \in [0, 1/2]$ is a risk aversion parameter associated with demand site i, $i \in I$, specified by decision makers according to their risk preferences. Let us denote the expected value of the random demand $h_i(\omega)$ by $\bar{h}_i$. We note that for the risk parameters $\alpha_i = 1/2$, $i \in I$, the ICCs (29) take the form of the coverage constraints, where the random variables are replaced by their expected values:

$$\sum_{j\in N_i} x_{ij} \ge \bar{h}_i, \quad i \in I.$$

Therefore, it is only meaningful to consider $\alpha_i \in \left[0, 1/2\right],\ i \in I$. Otherwise, we would obtain solutions for which the unmet demand values would be even higher than those associated with the solutions constructed by using a naive approach based on the expected demand values. Since $[w]_+ + [w]_- = |w|$ and $[w]_- = [w]_+ - w$ for $w \in \mathbb{R}$, the alternative ICCs (29) are equivalently represented by

$$(1-2\alpha_i)\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) \le \alpha_i \left(\sum_{j\in N_i} x_{ij} - \bar{h}_i\right), \quad i \in I. \tag{30}$$

Similarly, the ICC on the random total unmet demand defined in (27) is given by

$$(1-2\delta)\mathbb{E}\left(\sum_{i\in I}\left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) \le \delta \left(\sum_{i\in I}\sum_{j\in N_i} x_{ij} - \sum_{i\in I}\bar{h}_i\right), \tag{31}$$

where $\delta \in [0, 1/2]$ is a risk aversion parameter associated with the total unmet demand.

These constraints would allow the decision makers to evaluate different location and allocation decisions based on the tradeoff between the quality of service and costs by varying the risk parameters.

5.2 The optimization problem with integrated chance constraints

Replacing coverage constraints (2) in the *underlying deterministic problem* by ICCs (30) and (31) leads to the following stochastic programming problem:

$$\min \sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij} + \sum_{j\in J} f_j y_j + \sum_{i\in I}\sum_{j\in J} a x_{ij} \tag{32}$$

$$\text{subject to:} \quad (1-2\alpha_i)\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) \le \alpha_i \left(\sum_{j\in N_i} x_{ij} - \bar{h}_i\right), \quad \forall i \in I, \tag{33}$$

$$(1-2\delta)\sum_{i\in I}\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) \le \delta \left(\sum_{i\in I}\sum_{j\in N_i} x_{ij} - \sum_{i\in I}\bar{h}_i\right), \tag{34}$$

$$(\mathbf{x}, \mathbf{y}) \in Q, \tag{35}$$

where $\alpha_i \in [0, 1/2]$, $i \in I$, and $\delta \in [0, 1/2]$ are prescribed risk aversion parameters, and $Q = \{(\mathbf{x} \in \mathbb{Z}_+{}^{|I|\times|J|}, \mathbf{y} \in \{0,1\}^{|J|}) : \sum_{i\in M_j} x_{ij} \le U_j y_j,\ \forall j \in J\}$. Basically, we specify the risk constraints in (7) by ICCs (30) and (31). We refer to this problem as *the integrated chance constrained EMS system design* problem (ICCsP).

We set $\alpha_i, i \in I$, values to be equal for providing fair service to each demand site in terms of the proportion of the unmet demand. For the rest of the paper we let $\alpha_i = \alpha,\ i \in I$. The ICCs (33) defined for individual demand nodes are referred to as local constraints and the constraint (34) defined for the system-wide service level is referred to as a global constraint. We chose the risk parameters, $\delta < \alpha < 0.5$, so that both types of constraints drive the system. This model is significant since it allows us to control simultaneously the target levels for individual demand nodes and the entire network.

The optimization problem (32)–(35) can be represented by an MILP formulation by creating an $|S| \times |I|$ matrix of new variables representing the excess demand values. Recall that the excess value of demand realization $h_i^s,\ s \in S$, with respect to the total number of vehicles that are allocated due to the demand at node $i \in I$ is represented by $e_{(\mathbf{x},i)}(s)$ defined in (26). Then, we obtain the following deterministic equivalent formulation of ICCsP:

$$\min \sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij} + \sum_{j\in J} f_j y_j + \sum_{i\in I}\sum_{j\in J} a x_{ij} \tag{36}$$

$$\text{subject to:} \quad (1-2\alpha)\sum_{s\in S} p_s \hat{e}_{(\mathbf{x},i)}(s) \le \alpha \left(\sum_{j\in N_i} x_{ij} - \bar{h}_i \right), \quad \forall i \in I, \tag{37}$$

$$(1-2\delta)\sum_{i\in I}\sum_{s\in S} p_s \hat{e}_{(\mathbf{x},i)}(s) \le \delta \left(\sum_{i\in I}\sum_{j\in N_i} x_{ij} - \sum_{i\in I} \bar{h}_i \right), \tag{38}$$

$$\hat{e}_{(\mathbf{x},i)}(s) \ge h_i^s - \sum_{j\in N_i} x_{ij}, \quad \forall s \in S,\ i \in I, \tag{39}$$

$$\hat{e}_{(\mathbf{x},i)}(s) \ge 0, \quad \forall s \in S,\ i \in I, \tag{40}$$

$$(\mathbf{x}, \mathbf{y}) \in Q, \tag{41}$$

where decision variables $\hat{e}_{(\mathbf{x},i)}(s),\ s \in S,\ i \in I$, represent the excess values of demand realizations. We refer to this problem as DirectICCsP. The following observation is required to argue that we obtain the optimal solution even if we use variable $\hat{e}_{(\mathbf{x},i)}(s)$ instead of $\hat{e}_{(\mathbf{x},i)}(s),\ s \in S,\ i \in I$.

Observation 1 *For every feasible solution* $(\mathbf{x}, \mathbf{y}, \hat{e}_{(\mathbf{x},i)}(s),\ s \in S,\ i \in I)$ *of* (37)–(41) *we have* $\hat{e}_{(\mathbf{x},i)}(s) \ge e_{(\mathbf{x},i)}(s),\ i \in I$, *and the solution* $(\mathbf{x}, \mathbf{y}, e_{(\mathbf{x},i)}(s),\ s \in S,\ i \in I)$ *is also feasible for* (37)–(41).

One may solve this MILP formulation directly using a standard mixed integer programming solver such as CPLEX. However, in case of large instances it would be difficult for such a solver to provide an optimal solution. In the following section, we describe an alternative equivalent formulation for DirectICCsP in order to reduce the number of variables and develop an effective method to solve our original problem ICCsP.

5.3 An alternate formulation based on local demand distributions

A scenario represents a realization of joint service requests at the demand nodes. Notice that the same demand realization for a node can be observed under multiple scenarios, and therefore, the number of different demand realizations for each node would be significantly smaller than the number of global scenarios, $|S|$. For a given set of scenarios we can easily find the different demand realizations and the associated probabilities for each node. Basically, we decompose the global scenarios into local scenarios, and we denote the set of different demand realizations for node i by $\tilde{S}_i,\ i \in I$. Let $\tilde{p}_i(m)$ denote the probability that the realized value of the demand at node i is equal to $m,\ m \in \tilde{S}_i,\ i \in I$, and then $\tilde{p}_i(m) = \sum_{s\in S}\{p_s : h_i^s = m\}$ for all $i \in I,\ m \in \tilde{S}_i$. Then, the expected value of the random unmet demand at node $i \in I$ is rewritten as follows:

$$\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij} \right]_+ \right) = \sum_{s\in S} p_s \left[h_i^s - \sum_{j\in N_i} x_{ij} \right]_+ = \sum_{m\in \tilde{S}_i} \tilde{p}_i(m) \left[m - \sum_{j\in N_i} x_{ij} \right]_+ .$$

Introducing

$$\tau_{(\mathbf{x},i)}(m) = \left[m - \sum_{j \in N_i} x_{ij} \right]_+, \quad i \in I,\ m \in \tilde{S}_i,$$

the alternate deterministic equivalent formulation of ICCsP becomes

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j + \sum_{i \in I} \sum_{j \in J} a x_{ij}, \tag{42}$$

$$\text{subject to:} \quad (1-2\alpha) \sum_{m \in \tilde{S}_i} \tilde{p}_i(m) \hat{\tau}_{(\mathbf{x},i)}(m) \le \alpha \left(\sum_{j \in N_i} x_{ij} - \bar{h}_i \right), \quad \forall i \in I, \tag{43}$$

$$(1-2\delta) \sum_{i \in I} \sum_{m \in \tilde{S}_i} \tilde{p}_i(m) \hat{\tau}_{(\mathbf{x},i)}(m) \le \delta \left(\sum_{i \in I} \sum_{j \in N_i} x_{ij} - \sum_{i \in I} \bar{h}_i \right), \tag{44}$$

$$\hat{\tau}_{(\mathbf{x},i)}(m) \ge m - \sum_{j \in N_i} x_{ij}, \quad \forall i \in I,\ m \in \tilde{S}_i, \tag{45}$$

$$\hat{\tau}_{(\mathbf{x},i)}(m) \ge 0, \quad \forall i \in I,\ m \in \tilde{S}_i, \tag{46}$$

$$(\mathbf{x}, \mathbf{y}) \in Q, \tag{47}$$

where decision variables $\hat{\tau}_{(\mathbf{x},i)}(m),\ i \in I,\ m \in \tilde{S}_i$, are introduced to represent the excess values of demand realizations. We refer to this problem as *the alternative integrated chance constrained EMS system design* problem (AlterICCsP). This alternate formulation creates only $\sum_{i \in I} |\tilde{S}_i|$ new variables instead of $|S||I|$. Thus, it leads to a significant reduction in the number of variables and provides us with an effective solution method. For example, in our computational studies for a problem instance where $|I| = |J| = 400$ and $|S| = 50{,}000$ we have $\max_{i \in I} |\tilde{S}_i| = 9$ and $\sum_{i \in I} |\tilde{S}_i| = 2471$, whereas $|S||I| = 400 * 50{,}000 = 20{,}000{,}000$.

In the next section, we describe another model involving a different type of constraint on the system-wide service level, which is based on a common EMS performance criterion which imposes a lower bound on the fraction of calls whose response time is below a threshold (see Erkut et al. 2008a).

6 The stochastic dominance based EMS system design model

In Sect. 6.1, we discuss the SSD and increasing convex order (ICX) constraints. Then, in Sect. 6.2 we describe *the ICX-based EMS system design* problem and develop an equivalent alternate formulation. Since it is hard to solve the proposed problem with ICX constraints, we develop a heuristic procedure in Sect. 6.3. Finally, in Sect. 6.4 we describe how to utilize an EMS performance standard to generate the reference distribution of the total random unmet demand, which is required to apply the stochastic dominance based approach.

6.1 Stochastic dominance constraints

Comparing uncertain outcomes is one of the fundamental interests of decision theory. In many applications where the distribution of a random outcome is of significant interest, a single ICC may not be sufficient to model risk preferences. We may introduce many individual ICCs, which is closely related to the concept of SSD. The relation of stochastic dominance is one of the fundamental concepts of statistics and decision theory (Mann and Whitney 1947; Lehmann 1955). The concept of stochastic dominance introduces a preorder in the space of real random variables.[1] It has been widely used in economics and finance (e.g., see Hadar and Russell 1969; Levy 1992). We refer to Müller and Stoyan (2002) for a modern perspective on stochastic dominance relations. The SSD relation has been receiving significant attention due its correspondence with the risk-averse preferences. When larger values of random outcomes are preferred, we define the SSD relation based on the expected shortfall values as follows:

Definition 1 For two integrable random variables X and Y, X dominates Y in the second order, which we denote by $X \succeq_{(2)} Y$, if

$$\mathbb{E}([\eta - X]_+) \leq \mathbb{E}([\eta - Y]_+) \quad \text{for all } \eta \in \mathbb{R}. \tag{48}$$

The SSD relation corresponds to the "increasing convex order" rule which assumes the preference of smaller values to larger values. Thus, the increasing convex order relation can be used while comparing the random outcomes with the preference of smaller values and its definition is closely related to the one given above:

Definition 2 For two integrable random variables X and Y, X dominates (is stochastically smaller than) Y in "increasing convex order", which we denote by $X \preceq_{\text{icx}} Y$, if

$$\mathbb{E}([X - \eta]_+) \leq \mathbb{E}([Y - \eta]_+) \quad \text{for all } \eta \in \mathbb{R}. \tag{49}$$

In this study we focus on the random unmet demands, and therefore, we use the increasing convex order relation to compare random outcomes. However, note that equivalent formulations can be obtained by considering the random served demand (covered calls) and using the SSD relation. For the rest of paper, we focus on the ICX relation. Definition 2 featuring the expected excess values of the random variables is intuitive; preferring smaller realizations of a random outcome implies preferring smaller expected excess values with respect to some threshold values. Suppose that Y has a discrete distribution with realizations y_k, $k = 1, \ldots, D$, then the inequalities (49) are equivalent to (see Dentcheva and Ruszczyński 2003):

$$\mathbb{E}([X - y_k]_+) \leq \mathbb{E}([Y - y_k]_+), \quad k = 1, \ldots, D. \tag{50}$$

Stochastic dominance relations can be involved in optimization problems as constraints, allowing us to obtain solutions for which the random outcomes of interest dominate some benchmark (reference) random outcomes. Recently, there has been significant interest in such stochastic optimization models. They have been introduced and analyzed by Dentcheva and Ruszczyński (2003, 2006).

In this paper, we propose to introduce an ICX constraint into EMS system design models. In particular, we introduce a new type of EMS system design optimization model involving

[1] A preorder is a relation that is reflexive and transitive, but not necessarily antisymmetric and each preorder induces an equivalence relation between elements.

local ICCs on random unmet demands and a single ICX constraint on the random total unmet demand. The ICX constraint allows us to construct location and allocation decisions for which the random total unmet demand dominates a reference random total unmet demand. In this study, we define the reference outcome based on a common EMS performance standard which is to respond to at least $\rho * 100\%$ of all calls within r_1 minutes. Therefore, our model with the ICX constraint constructs solutions consistent with this common EMS performance standard. By proposing the second type of stochastic optimization model, we intend to provide a useful analytical tool for the EMS decision makers who are more interested in the distribution of the random outcome and would like to obtain a decision vector dominating a potential one under consideration.

6.2 The optimization problem with increasing convex order constraints

Consider our problem of designing an EMS system in which the decision vector $\mathbf{x}$ affects random unmet demands, $e_{(\mathbf{x},i)}, i \in I$, defined by (26). In order to find the best feasible decision vector $\mathbf{x}$, we compare the corresponding random unmet demands according to a preference relation. Here, we specify the preference relation among random variables based on ICCs (33) and the increasing convex order relation. Let Y be a random variable representing a *benchmark random total unmet demand*. We develop the following stochastic optimization model involving ICCs and an ICX constraint:

$$\min \sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij} + \sum_{j\in J} f_j y_j + \sum_{i\in I}\sum_{j\in J} a x_{ij},$$

$$\text{subject to:} \quad (1-2\alpha)\mathbb{E}\left(\left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) \le \alpha\left(\sum_{j\in N_i} x_{ij} - \bar{h}_i\right), \quad \forall i \in I, \tag{51}$$

$$\xi_{(\mathbf{x})} \preceq_{\text{icx}} Y, \tag{52}$$

$$(\mathbf{x},\mathbf{y}) \in Q. \tag{53}$$

We refer to this problem as *the ICX-based EMS system design* problem (ICXP).

Note that, equivalently, one can consider the SSD relation for the random total served demand instead of the ICX relation for the random total unmet demand. First, observe that the random total served demand is equal to

$$\sum_{i\in I} \min\left(h_i(\omega), \sum_{j\in N_i} x_{ij}\right) = \sum_{i\in I}\left(h_i(\omega) - \left[h_i(\omega) - \sum_{j\in N_i} x_{ij}\right]_+\right) = \sum_{i\in I} h_i(\omega) - \xi_{(\mathbf{x})}.$$

Then by Definitions 1 and 2, it is easy to see that

$$\sum_{i\in I} h_i(\omega) - \xi_{(\mathbf{x})} \succeq_{(2)} \sum_{i\in I} h_i(\omega) - Y \quad \Leftrightarrow \quad \xi_{(\mathbf{x})} \preceq_{\text{icx}} Y.$$

Thus, an equivalent formulation of ICXP based on an SSD constraint can be obtained by replacing $\xi_{(\mathbf{x})} \preceq_{\text{icx}} Y$ by $-\xi_{(\mathbf{x})} \succeq_{(2)} -Y$.

Suppose that the reference random outcome Y has a discrete distribution with realizations $y_1 < y_2 < \cdots < y_D$. Then, by the use of (50) the ICX relation (52) is equivalent to

$$\mathbb{E}([\xi_{(\mathbf{x})} - y_k]_+) \le \mathbb{E}([Y - y_k]_+), \quad k = 1, \ldots, D. \tag{54}$$

We would like to point out that the set of inequalities (54) can be viewed as finitely many ICCs of type (28). In order to solve problem ICXP effectively we propose to use the set of linear inequalities introduced by Luedtke (2008) for the SSD relation, where larger realizations of an uncertain outcome are preferred. We adapt this set of inequalities to the case of reverse preference, since the smaller values of the total unmet demand are preferred.

Theorem 1 (A representation of the ICX relation) *Suppose that X is a random variable and the realization of X under scenario s is denoted by $z_s, s \in S$. Furthermore, let Y be a discrete distribution with the realizations y_k, $k = 1, \dots, D$, and v_k, $k = 1, \dots, D$, denote the associated probabilities. Then, $X \preceq_{\text{icx}} Y$ if and only if there exists $\pi \in \mathbb{R}_+^{|S|D}$ such that*

$$z_s \leq \sum_{k=1}^{D} y_k \pi_{sk}, \quad \forall s \in S, \tag{55}$$

$$\sum_{k=1}^{D} \pi_{sk} = 1, \quad \forall s \in S, \tag{56}$$

$$\sum_{s \in S} \sum_{j=k}^{D} (y_j - y_k)\pi_{sj} \leq \sum_{j=k}^{D} v_j (y_j - y_k), \quad k = 1, \dots, D. \tag{57}$$

In our study, z_s is the realization of the total unmet demand under scenario s, i.e., $z_s = \xi_{(\mathbf{x})}(s) = \sum_{i \in I} e_{(\mathbf{x},i)}(s)$. Here, we cannot utilize the local demand distributions as we did for ICCsP, since we additionally consider the total unmet demand, and thus, the global scenarios cannot be decomposed. Then, using the variables representing the excess demand values we obtain an MILP formulation of ICXP referred to as AlterICXP:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j + \sum_{i \in I} \sum_{j \in J} a x_{ij}$$

$$\text{subject to:} \quad (1 - 2\alpha) \sum_{s \in S} p_s \hat{e}_{(\mathbf{x},i)}(s) \leq \alpha \left(\sum_{j \in N_i} x_{ij} - \bar{h}_i \right), \quad \forall i \in I, \tag{58}$$

$$(39)–(40), \tag{59}$$

$$\sum_{i \in I} \hat{e}_{(\mathbf{x},i)}(s) \leq \sum_{k=1}^{D} y_k \pi_{sk}, \quad \forall s \in S, \tag{60}$$

$$(56)–(57), \tag{61}$$

$$\pi_{sk} \geq 0, \quad \forall s \in S, \ k = 1, \dots, D, \tag{62}$$

$$(\mathbf{x}, \mathbf{y}) \in Q. \tag{63}$$

This formulation creates $|S| * D$ new variables and $O(S + D)$ constraints in order to represent the ICX relation (52) by linear inequalities. To improve the computational effectiveness we propose to rewrite ICCs (58) using a binary search algorithm. Notice that ICCs (51) are local constraints, each of which independently imposes a lower bound on the total number of vehicles to be allocated in order to ensure the individual target service levels. It is easy to see that these constraints depend monotonically on the number of vehicles allocated to cover demand at each node, i.e., if $\sum_{j\in N_i} x_{ij} = k$ satisfies the ICC associated with node i, then for any number of vehicles greater than k the same constraint is satisfied. Due to this special structure of the local constraints, we implement a binary search algorithm for each demand node in order to find the minimum number of vehicles needed to reach the individual target service levels. We denote these lower bounds on the number of vehicles allocated due to the demand at each node by $LB_i^{\text{ICC}},\ i \in I$. Finally, the proposed formulation for ICXP, which we refer to as AlterICXPWithBS, is given by

$$\begin{aligned} \min \quad & \sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij} + \sum_{j\in J} f_j y_j + \sum_{i\in I}\sum_{j\in J} a x_{ij} \\ \text{subject to:} \quad & \sum_{j\in N_i} x_{ij} \geq LB_i^{\text{ICC}}, \quad \forall i \in I, \\ & (59)-(63). \end{aligned}$$

6.3 Heuristic algorithm

In our numerical study, we could solve moderately large problem instances using the proposed formulation AlterICXPWithBS. However, solving AlterICXPWithBS requires substantially more effort as the number of potential sites and the number of scenarios increase due to the variables introduced to represent the ICX relation. Hence, we propose a mathematical programming based heuristic procedure, HICXP, that yields a feasible solution with a small optimality gap within a reasonable computation time. Intuitively, we have a two step problem. First, we need to find an estimate for the optimal number of vehicles required to satisfy the target service levels, which is accomplished by solving the LP relaxation of

Algorithm 1 Heuristic algorithm HICXP for ICXP

Initialize the iteration number, $l = 0$.
Solve the LP relaxation of AlterICXPWithBS to obtain the optimal solution $(\mathbf{x}^{LP}, \mathbf{y}^{LP})$.
Assign the total number of vehicles allocated due to node i, $h_i^l, \forall i \in I$:

$$h_i^l = \begin{cases} \lfloor \sum_{j\in N_i} x_{ij}^{LP} \rfloor, & \text{if } \lceil \sum_{j\in N_i} x_{ij}^{LP} \rceil - \sum_{j\in N_i} x_{ij}^{LP} > 0.5, \\ \lceil \sum_{j\in N_i} x_{ij}^{LP} \rceil, & \text{otherwise.} \end{cases}$$

if ICX constraints (54) are not satisfied for $\sum_{j\in N_i} x_{ij} = h_i^l$, $i \in I$, **then**
    Find the set $I^* = \{i \in I : (\lceil \sum_{j\in N_i} x_{ij}^{LP} \rceil - h_i^l) > 0\}$.
end if
while ICX constraints (54) are not satisfied for $\sum_{j\in N_i} x_{ij} = h_i^l$, $i \in I$, **do** {Iteratively increase the number of vehicles allocated}
    Let $l := l + 1$. Find the smallest index $i^* \in I^*$ such that $i^* = \arg\max_{i\in I^*}\{\lceil \sum_{j\in N_i} x_{ij}^{LP} \rceil - h_i^l\}^{l-1}$.
    Update the allocation vector $\mathbf{h}^l : h_{i^*}^l = \lceil \sum_{j\in N_{i^*}} x_{i^*j}^{LP} \rceil$ and $h_i^l = h_i^{l-1}$ for $i \in I \setminus i^*$.
    Let $I^* := I^* \setminus i^*$.
end while
Solve the *the underlying deterministic problem* defined by (1)–(5), where $h_i = h_i^l,\ i \in I$.

AlterICXPWithBS. Then, given the vehicle requirements we solve the deterministic problem (1)–(5) in order to determine the facility locations and the vehicle allocation at each facility.

The finite heuristic algorithm, described in Algorithm 1, provides us with a feasible solution of ICXP. It is easy to see that $\sum_{j\in N_i} x_{ij} = \lceil \sum_{j\in N_i} x_{ij}^{LP} \rceil$ would satisfy the ICX constraints. However, such a direct round up approach would lead to an excessive allocation of vehicles. Thus, instead we propose to increase the number of vehicles to be allocated iteratively until the feasibility of ICX constraints is achieved. This iterative procedure turns out to be effective to obtain feasible solutions with reasonable small optimality gaps (see Sect. 7.4).

6.4 How to generate the distributions of reference outcomes

As discussed before, applying the stochastic dominance based approach requires that the reference distribution of the total random unmet demand is available in advance. To this end, we need to specify the target values $y_k,\ k = 1, \dots, D$, as realizations of the reference total unmet demand along with the associated probabilities $v_k,\ k = 1, \dots, D$.

A widely applied EMS standard is to respond to $\rho * 100\%$ of all calls within r_1 minutes (generally, $\rho = 0.9$ and $r_1 = 8$). In order to construct solutions in line with this standard we construct the empirical distribution of the total demand, denoted by TD, given the set of scenarios and their associated probabilities. Then, we set the reference random total unmet demand to $Y = (1 - \rho)\,\mathrm{TD}$. For instance, suppose that a total demand realization is 50 with probability 0.001. For $\rho = 0.9$, we have $50 * 0.1 = 5$ as a reference target value for the total unmet demand with the probability of 0.001. The parameter $r_1 = 8$ is taken into consideration while defining the set of demand nodes that can be covered by a facility located at node $j \in J$ and the set of all candidate facility nodes that are within acceptable distance of node $i \in I$. For illustrative examples, please see Figs. 1(a) and 1(b).

The parameter values $\rho = 0.9$ and $r_1 = 8$ are common in North America, but not necessarily in elsewhere in the world. However, the method described above can be applied for any parameter values. Note also that alternatively a reference total unmet demand may be defined based on a potential solution. For example, we can solve a simplified deterministic version of the EMS design problem and obtain its optimal solution. For that solution, we can calculate the unmet demands under each scenario and obtain the empirical distribution of the random total unmet demand. Then, this distribution can be used to define a benchmark distribution according to the decision maker's risk preference.

7 Computational results

In the following section, we give some details on generating the problem instances. Then, in Sect. 7.2 we present illustrative results to give some insights about the level of increase in the total cost when the single-stage versions of the proposed models are solved instead of the two-stage versions. In Sect. 7.3, we provide results to demonstrate the computational effectiveness of the proposed alternate formulations of ICCsP and ICXP. Section 7.4 presents numerical results illustrating the computational effectiveness of the heuristic developed for ICXP. In Sects. 7.5 and 7.6, we present numerical results to analyze how the optimal location and allocation solutions change with respect to the input parameters and different risk preferences represented by ICCs and the ICX constraint. Finally, we discuss the computational study performed to compare the proposed models to the most relevant existing model (Beraldi et al. 2004).

7.1 Generation of problem instances

In order to test the computational performance of our solution methods, we consider several problem instances of different sizes. A total of 206 test problem instances are used to obtain the numerical results presented in this section. For a specified number of nodes ($n = |J| = |I|$) in the network, where each node is a demand point as well as a facility candidate, problem instances were randomly generated as follows:

- We randomly generate the set of demand points in the $[0, 30]^2$ square according to a continuous uniform distribution as proposed by Gendreau et al. (1997). We set d_{ij}, $i \in I$, $j \in J$, values to be the Euclidean distance between these points. As Gendreau et al. (1997), we assume the side of the square region to be 30 km and the ambulance speed to be 40 km/h. Then, the coverage distance threshold T_c is $40 * 8/60 \approx 5.33$ km, when the response time standard is chosen to be 8 minutes.
- The (annual) fixed facility cost vector $365 * 24 * \mathbf{f}$ is sampled from the uniform distribution on the interval $[1000, 4000]$. The (annual) cost of purchasing and maintaining an emergency vehicle ($365 * 24 * a$) is 100. The cost per unit distance per unit demand (β) is 0.001 or 0.01.
- Demand realizations at each node $i \in I$, h_i^s, $s \in S$, are generated from a Poisson distribution with the arrival rate parameter λ_i, which is sampled from the uniform distribution on the interval $[0.1, 0.8]$. As an alternative, in order to allow higher demand density in the city center, we divide the $[0, 30]^2$ square in 9 equal square zones. The arrival rate parameter for the nodes in the city center is sampled from the uniform distribution on the interval $[0.8, 1]$. For the corner zones we use the interval $[0.1, 0.2]$, whereas, for the remaining four zones we use the interval $[0.3, 0.4]$.
- The maximum number of vehicles that can be allocated to each facility $j \in J$, U_j, is sampled from the uniform distribution on the interval $[2, 4]$ or $[6, 10]$.
- Scenario probabilities p_s, $s \in S$, are set to be equal or sampled from the uniform distribution on the interval $[0.2, 0.7]$ and then normalized.
- The risk aversion parameter α for individual ICCs is chosen to be 0.2, whereas the risk aversion parameter δ for the global ICC is chosen to be 0.04 or 0.02.
- The proportion of demand covered within $r_1 = 8$ minutes (ρ) is 0.9 unless stated otherwise. As discussed in Sect. 6.4, the ρ parameter is used to generate the reference random total unmet demand.

We would like to point out that generating the scenarios is not our main concern here. Existing methods can be applied to generate alternate scenarios, or if available, real historical data may be employed. Also note that the demand varying depending on the time of day or the day of the week, can be incorporated into the model by generating demand realizations for different hourly time periods of a week.

All problems were solved using the AMPL modeling language (Fourer et al. 1993) running on ILOG CPLEX 10.2 (ILOG 2006). The binary search algorithms were implemented in MATLAB R2006. The numerical experiments were performed on a 64-bit, 2 quad-core CPU HP workstation running on Linux. In our computational study, we terminated CPLEX when the prescribed time limit of 7200 seconds is reached.

In order to give an idea about the cost structure in the generated data sets, we present in Table 1 the average optimal objective function values decomposed into three types of cost components for 23 test problems presented in Table 3. We assume that the cost associated with the amount of labor needed to operate facilities is also incorporated into the fixed costs $f_i, i \in I$, associated with the facility locations. As labor is typically the dominating cost

Table 1 Average optimal cost values associated with Table 3

Variable cost $\sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij}$	Fixed cost $\sum_{j\in J} f_j y_j$	Vehicle cost $\sum_{i\in I}\sum_{j\in J} a x_{ij}$	Total cost
6,880.57	57,096.65	36,752.17	100,729.39
6.83%	56.68%	36.49%	100.00%

Table 2 Results for the proposed single- and two-stage formulations

Family of prob. instances	$\sum_{i\in I}\lvert N_i\rvert$	$\sum_{i\in I}\lvert N_i\rvert * \lvert S\rvert$	With local ICCs and global ICC constraint			With local ICCs and global ICX constraint		
			Total cost	CPU		Total cost	CPU	
			Percentage increase	Two-stage	Single-stage	Percentage increase	Two-stage	Single-stage
$n=25$ $\lvert S\rvert=50$	78	3900	7.53%	32.39	0.11	10.63%	330.93	0.06
$n=25$ $\lvert S\rvert=100$	85.4	8540	7.04%	321.75	0.05	16.00%	19.00	0.13
$n=25$ $\lvert S\rvert=200$	81.4	16280	6.02%	1372.40	0.18	18.28%	84.27	0.21
$n=25$ $\lvert S\rvert=300$	75	22500	5.77%	2032.06	0.05	19.26%	91.11	0.30
$n=50$ $\lvert S\rvert=50$	258	12900	13.69%	1255.18	0.52	20.44%	179.97	0.33
$n=50$ $\lvert S\rvert=200$	256	51200	<16.03%*	7199.63*	0.22	24.97%	1173.55	1.24

*For all 5 instances of the two-stage formulation, CPLEX terminated due to the time limit

component (Goldberg 2004), the cost parameters are chosen accordingly in order to give the largest priority to the fixed costs of facilities in our models.

7.2 On safe approximation

Recall that the single-stage formulations are based on the simplifying assumption that allocations do not depend on scenarios. As discussed in Sect. 4, the risk constraints in the single-stage formulations are safe approximations of those in the two-stage formulations. Here, we present numerical results to demonstrate that the two-stage models would provide decisions with lower total costs.

We cannot directly solve the deterministic equivalent formulations of the two-stage versions of the proposed models for moderate size problem instances. Therefore, we present comparative results for several small problem instances. Let Obf^1 and Obf^2 denote the optimal objective function values (total costs) of the single- and two-stage versions of the proposed models (ICCsP or ICXP), respectively. Then we calculate the percentage increase (PI) in the total cost as follows:

$$PI = \frac{\text{Obf}^1 - \text{Obf}^2}{\text{Obf}^2}. \tag{64}$$

Table 2 shows the percentage increases in the total cost, averaged over 5 problem instances for each family. Even for these small problem instances, CPLEX could not find all of the optimal solutions for the two-stage versions of the models within the time limit. For such instances, the percentage increase is calculated by replacing Obf^2 in (64) with the best lower bound on the total cost provided by CPLEX.

There is definitely a tradeoff between the computational complexity and the solution quality. The main focus of this paper is to introduce alternate risk measures in stochastic EMS design models and deal with large problem instances, and due to the computational challenge of the two-stage formulations, we dedicate our study to the single-stage models. In the rest of the section, we report results only on the single-stage formulations.

7.3 On alternate formulations

Here, we present some results indicating how substantial the computational improvement is after decomposing the global scenarios into the local ones. Table 3 shows the CPU times that are in favor of AlterICCsP even for the small problem instances.

It can be seen from Table 4 that when the formulation AlterICXP is used, CPLEX could not provide optimal solutions within the time limit of 7200 seconds even for small instances of ICXP. However, these instances could be solved to optimality by using the formulation AlterICXPWithBS. The proposed formulation AlterICXPWithBS utilizing binary search algorithms substantially outperforms the formulation AlterICXP and leads to great reductions

Table 3 Performance of AlterICCsP versus DirectICCsP

Problem instances	$\sum_{i\in I} \tilde{S}_i$	DirectICCsP CPU (sec.)	AlterICCsP CPU (sec.)	Relative reduction in CPU (%)
$n = 100$ $\|S\| = 500$	430	125.96	1.24	99.02%
$n = 150$ $\|S\| = 500$	652	852.15	7.68	99.10%
$n = 200$ $\|S\| = 500$	884	1780.02	23.78	98.66%
$n = 300$ $\|S\| = 500$	1326	1692.14	25.17	98.51%
$n = 400$ $\|S\| = 500$	1784	5924.99	94.08	98.41%
$n = 100$ $\|S\| = 1000$	459	2928.00	3.94	99.87%
$n = 150$ $\|S\| = 1000$	704	1134.38	4.34	99.62%
$n = 200$ $\|S\| = 1000$	941	7210.96*	19.34	>99.73%
$n = 300$ $\|S\| = 1000$	1383	7219.09*	20.58	>99.71%
$n = 400$ $\|S\| = 1000$	1904	7235.69*	44.10	>99.39%
$n = 100$ $\|S\| = 3000$	507	7210.17*	6.14	>99.91%
$n = 150$ $\|S\| = 3000$	772	7218.36*	30.82	>99.57%
$n = 200$ $\|S\| = 3000$	1011	7230.04*	70.83	>99.02%
$n = 300$ $\|S\| = 3000$	1556	7617.45*	109.67	>98.56%
$n = 400$ $\|S\| = 3000$	2073	7295.23*	184.85	>97.47%
$n = 300$ $\|S\| = 10000$	1671	N/A	34.29	N/A
$n = 400$ $\|S\| = 10000$	2230	N/A	426.12	N/A
$n = 200$ $\|S\| = 30000$	1205	N/A	29.25	N/A
$n = 300$ $\|S\| = 30000$	1841	N/A	47.91	N/A
$n = 400$ $\|S\| = 30000$	2409	N/A	67.11	N/A
$n = 200$ $\|S\| = 50000$	1241	N/A	18.36	N/A
$n = 300$ $\|S\| = 50000$	1906	N/A	40.23	N/A
$n = 400$ $\|S\| = 50000$	2471	N/A	170.89	N/A

*Time limit with integer solution

N/A: No solution is available since CPLEX terminated due to solver error (ran out of memory)

Table 4 Performance of AlterICXPWithBS versus AlterICXP

Problem instances	AlterICXP CPU (sec.)	AlterICXPWithBS CPU (sec.)	Relative Reduction in CPU (%)
$n = 50$ $\|S\| = 300$	76.06	11.60	84.75%
$n = 75$ $\|S\| = 300$	118.14	14.42	87.80%
$n = 100$ $\|S\| = 300$	2129.35	82.21	96.14%
$n = 150$ $\|S\| = 300$	7201.81*	137.52	>98.09%
$n = 200$ $\|S\| = 300$	7202.95*	1043.47	>85.51%
$n = 50$ $\|S\| = 500$	43.07	5.33	87.62%
$n = 75$ $\|S\| = 500$	754.54	84.85	88.75%
$n = 100$ $\|S\| = 500$	7201.29*	1213.52	>83.15%
$n = 150$ $\|S\| = 500$	3692.95	442.86	88.01%

*CPLEX terminated due to the time limit

in CPU time to solve ICXP. We remark that the CPU time to obtain the local sets $\tilde{S}_i,\ i \in I$, and the associated probabilities from a given set of global scenarios (S), and to calculate the lower bounds $LB_i^{\mathrm{ICC}},\ i \in I$, were negligible and therefore, we did not report them. For example, for a large problem instance with $n = 400$ and $|S| = 50000$, we obtain the sets $\tilde{S}_i,\ i \in I$, and the associated probabilities using MATLAB R2006 in 6.02 CPU seconds. For the same problem instance it took the binary search algorithms 0.06 CPU seconds in total to calculate the lower bounds, $LB_i^{\mathrm{ICC}}, i \in I$.

7.4 On the heuristic HICXP

Results presented in this section illustrate the computational effectiveness of the heuristic HICXP. Since it is hard to solve AlterICXPWithBS for large problem instances, we solve smaller problem instances in order to calculate the upper bounds on the optimality gaps, as defined below. We use the best known lower bound on the objective value found by the branch-and-bound algorithm of CPLEX, as most of the problem instances cannot be solved for optimality within the prescribed time limit. Let Obf_t denote the best lower bound on the objective function value that is provided by the CPLEX solver, when the prescribed time limit t is reached. Obf^* denotes the objective function value obtained by HICXP. The feasible solution obtained by the described heuristic algorithm gives an upper bound on the objective value. Then, we define the upper bound on the optimality gap (UBOPG) associated with the heuristic HICXP as follows:

$$\mathrm{UBOPG} = \frac{\mathrm{Obf}^* - \mathrm{Obf}_t}{\mathrm{Obf}_t}.$$

To take the randomness in data generation into account, 5 test problem instances were generated for each selected combination of parameters n and $|S|$. Table 5 reports the CPU times and UBOPG values averaged over 5 randomly generated instances. This table shows that heuristic HICXP is quite effective for generating feasible solutions with reasonably small optimality gaps.

Numerical results presented in Sects. 7.3 and 7.4 indicate that the proposed solution methods are quite sufficient to solve ICCsP and ICXP for large problem instances.

Table 5 Performance of Heuristic HICXP

Family of problem instances	ICXP CPU (sec.)	HICXP CPU (sec.)	Relative reduction in CPU (%)	UBOPG
$n = 100$ $\lvert S\rvert = 300$	55.27	8.35	84.90%	2.28%
$n = 150$ $\lvert S\rvert = 300$	378.45	18.87	95.01%	1.49%
$n = 200$ $\lvert S\rvert = 300$	3839.03*	30.31	>99.21%	0.75%
$n = 300$ $\lvert S\rvert = 300$	7219.26*	87.69	>98.79%	0.74%
$n = 400$ $\lvert S\rvert = 300$	7209.33*	223.47	>96.90%	0.76%
$n = 100$ $\lvert S\rvert = 500$	2008.75*	18.13	>99.10%	3.46%
$n = 150$ $\lvert S\rvert = 500$	4844.03*	50.44	>98.96%	1.72%
$n = 200$ $\lvert S\rvert = 500$	7209.28*	75.31	>98.96%	1.23%
$n = 300$ $\lvert S\rvert = 500$	7216.89*	233.30	>96.77%	0.84%
$n = 400$ $\lvert S\rvert = 500$	7219.07*	469.80	>93.49%	0.56%
$n = 100$ $\lvert S\rvert = 1000$	5685.66*	54.41	>99.04%	2.91%
$n = 150$ $\lvert S\rvert = 1000$	6866.01*	155.04	>97.74%	2.43%
$n = 200$ $\lvert S\rvert = 1000$	7207.28*	304.96	>95.77%	1.89%
$n = 300$ $\lvert S\rvert = 1000$	7215.96*	889.93	>87.67%	1.09%
$n = 400$ $\lvert S\rvert = 1000$	7235.78*	1875.20	>74.08%	0.98%

*CPLEX terminated due to the time limit (for at least one instance out of 5)

Table 6 Results for different input parameters (averaged over 5 test problems)

	Total cost	Total # of facilities	Ratio of total # of vehicles	Total expected unmet demand
Base case	79,488.66	33.4	0.4233	8.39
Response time standard: 7 mins.	81,098.91	34.0	0.4239	8.41
Upper bounds U_i, $i \in I$, divided by 2	146,279.19	70.4	0.4233	8.39
Upper bounds U_i, $i \in I$, multiplied by 2	60,204.50	18.5	0.4241	8.42
Risk parameter $\delta = 0.01$	110,141.35	45.4	0.5731	2.99
Risk parameter $\delta = 0.1$	74,968.00	31.8	0.3979	10.08
Fixed costs, f_i, $i \in I$, divided by 20	35,635.77	40.0	0.4240	8.40
Fixed costs, f_i, $i \in I$, multiplied by 100	4,498,674.53	33.2	0.4242	8.41
Unit purchase cost, a, divided by 1000	51,113.48	33.4	0.4251	8.45
Unit distance cost, β, multiplied by 100	429,547.36	67.8	0.4380	8.82

7.5 Solution sensitivity to input parameters

We solve ICCsP for a particular problem instance with $n = 200$ and $|S| = 500$ in order to check the sensitivity of the model to changes in input parameters. We refer to the originally generated test problem instance as the *base case*. The value(s) of only a certain type of parameter is (are) changed while everything else is kept fixed. The entries in the first column of Table 6 state which parameter values have been changed in the data of the *base case*. As expected the upper bounds on the number of vehicles that can be allocated to facilities have a significant effect on the number of facilities to be opened. It can also be seen from Table 6 that the risk parameter δ significantly effects the number of vehicles to be allocated. We will elaborate on the effects of the risk parameters more in the next section.

The fixed cost parameters of the *base case* are turned out to be large enough to construct solutions with the main objective of minimizing the number of facilities to be opened. That is why when we increase the values of parameters $f_i, i \in I$, the optimal solution does not really change. The results also show that when we decrease the values of the fixed costs $f_i, i \in I$, the number of vehicles allocated stays almost constant, whereas the number of facilities to be opened increases significantly. When the unit variable cost β is set to be a high value the model may result in more facilities to reduce the total cost. However, we prefer small values for the unit variable cost in order to give lower priority to the total variable cost than the other cost components. Due to the risk constraints imposing lower bounds on the number of vehicles, decreasing the value of the cost parameter a does not have a significant effect on the total number of vehicles allocated and the number of facilities to be opened. Note also that by the same reasoning, the total number of vehicles allocated and the total expected unmet demand are mainly affected by the changes in the risk parameters.

Using a worst case scenario approach, a decision maker might construct a solution based on the highest demand realizations to avoid any unmet demand situations, i.e., demand would be satisfied under all scenarios. Instead of the number of vehicles, we report the ratio of the total number of vehicles allocated to the total number vehicles that would be required based on the worst case scenario approach. We will refer to this ratio as the "ratio of total number of vehicles".

7.6 On risk constraints

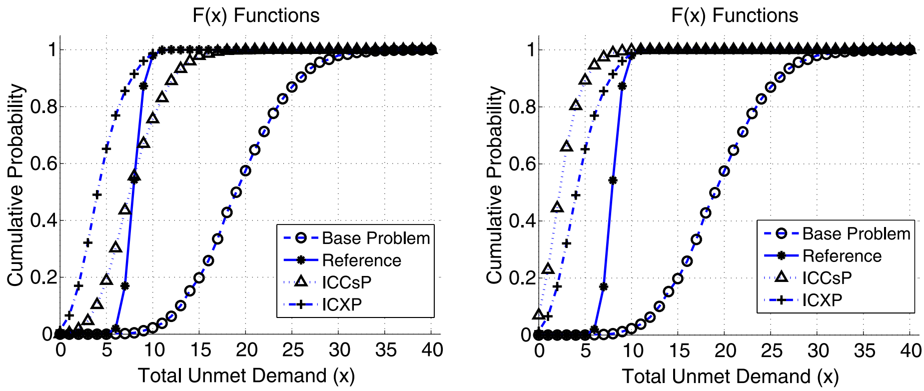
For comparison purposes, we find optimal decisions by solving a problem based on a naive approach which uses expected demand values. We refer to the *underlying deterministic problem* (1)–(5), where h_i set to $\lceil \bar{h}_i \rceil$, $\forall i \in I$, as the *base problem*. The stochastic models, which consider the variability in demand, provide solutions with more flexibility to avoid unmet demand situations. Therefore, it is not surprising that the solutions of our models are better in terms of the amount of total unmet demand than the ones found by the base problem (see Table 7). As it can be seen from Table 7, the optimal numbers of vehicles obtained by ICCsP and ICXP are between the values obtained by the base problem and the worst case scenario approach (ratio of total # of vehicles = 1). Thus, risk measures deal with the problem of placing too much weight on the extreme scenarios.

Table 7 Total number of facilities, ratio of total # of vehicles and expected unmet demand by the base problem, ICCsP and ICXP

Problem instances	Total number of facilities			Ratio of total # of vehicles			Total expected unmet demand		
	Base prob.	ICCsP	ICXP	Base prob.	ICCsP	ICXP	Base prob.	ICCsP	ICXP
$n = 100$ $\lvert S\rvert = 3000$	16	18	23	0.25	0.35	0.47	9.19	4.23	1.64
$n = 200$ $\lvert S\rvert = 3000$	24	33	41	0.25	0.34	0.42	19.39	8.28	4.50
$n = 300$ $\lvert S\rvert = 3000$	34	50	58	0.24	0.34	0.40	32.18	12.80	8.80
$n = 100$ $\lvert S\rvert = 5000$	15	17	24	0.23	0.33	0.44	9.80	4.30	1.57
$n = 150$ $\lvert S\rvert = 5000$	18	25	32	0.22	0.29	0.41	14.82	6.40	3.02
$n = 200$ $\lvert S\rvert = 5000$	23	32	40	0.23	0.32	0.40	18.93	8.27	4.35

Table 8 For an instance with $n = 200$ and $|S| = 3000$

Risk parameter	For reference benchmark	Ratio of total # of vehicles		Total # of facilities	
δ	ρ	ICCsP	ICXP	ICCsP	ICXP
0.04	0.9	0.34	**0.41**	31	**38**
0.01	0.9	**0.46**	0.41	**43**	38



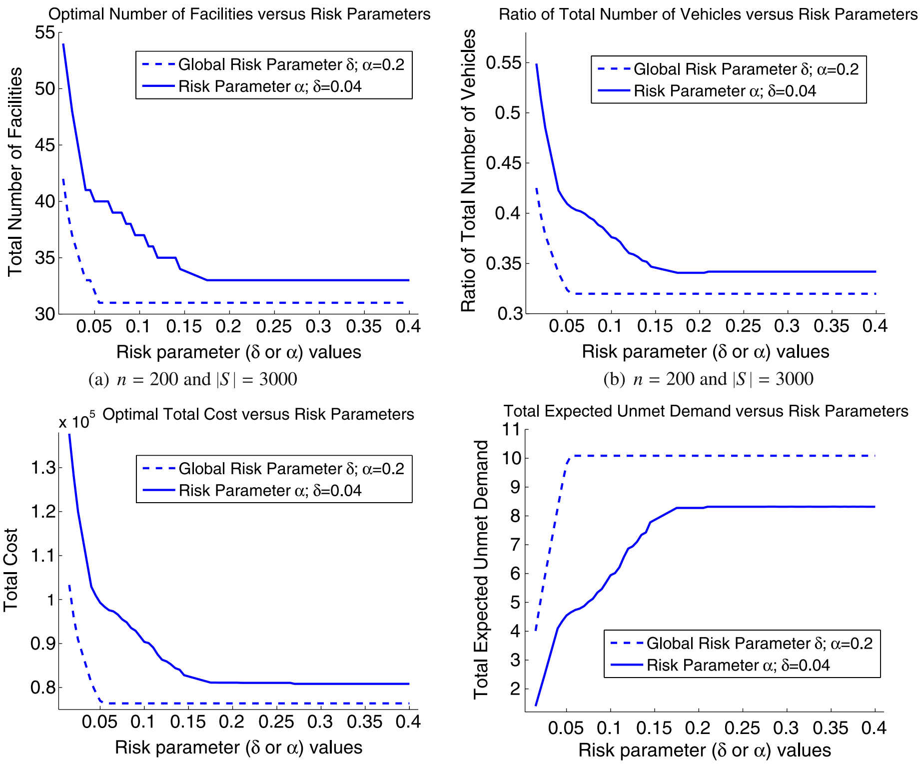
(a) $n = 200$, $|S| = 3000$, $\delta = 0.04$ (b) $n = 200$, $|S| = 3000$, $\delta = 0.01$

Fig. 1 Reference distribution function and distribution functions of the total unmet demand associated with the optimal solutions of ICCsP, ICXP and the base problem

The level of risk modeled by ICCsP and ICXP depend on the risk parameters and the specified distribution of the reference random outcome. Either of these two types of models may provide us with more risk averse solutions according to the risk parameter δ and the reference distribution. In order to compare the level of risk aversion of the optimal solutions we can check the distribution of total unmet demand when exactly the optimal number of vehicles are assigned to each demand node, and the distribution of reference random total unmet demand. For example, for a problem instance where $n = 200$ and $|S| = 3000$, the optimal total number of vehicles required according to ICCsP is smaller than the amount required according to ICXP when $\delta = 0.04$ (see Table 8). For this instance, by setting the risk parameter δ to 0.04 and 0.01, and keeping everything else fixed, we obtain two different Figs. 1(a) and 1(b). These figures show the reference cumulative distribution function and cumulative distribution functions of the random total unmet demand associated with the optimal solutions of ICCsP, ICXP and the base problem. We note that when $\delta = 0.04$ the total unmet demand for the optimal solution constructed by ICCsP is dominated by the reference random total unmet demand in the increasing convex order. Therefore, ICXP which requires the ICX dominance with respect to the reference random outcome leads to a more risk averse solution than the one obtained by ICCsP. On the other hand, when $\delta = 0.01$ the reference random total unmet demand is dominated by the total unmet demand for the optimal solution obtained by ICCsP in the increasing convex order. Therefore, in this case the ICCsP results in a more risk averse solution; the total numbers of vehicles assigned to cover demand is higher (comparing ratios: $0.46 > 0.41$). In conclusion, we cannot claim that ICCsP provides more risk averse solutions than ICXP for all values of the risk parameter δ. It can also be seen from the results in Table 9 corresponding to a problem instance, where

Table 9 For an instance with $n = 200$ and $|S| = 500$

Risk parameter	For reference benchmark	Ratio of total # of vehicles		Total # of facilities	
δ	ρ	ICCsP	ICXP	ICCsP	ICXP
0.04	0.85	**0.42**	0.41	**34**	33
0.04	0.9	0.42	**0.48**	34	**39**



(a) $n = 200$ and $|S| = 3000$ (b) $n = 200$ and $|S| = 3000$

(c) $n = 200$ and $|S| = 3000$ (d) $n = 200$ and $|S| = 3000$

Fig. 2 Results obtained solving ICCsP for different values of risk aversion parameters

$n = 200$ and $|S| = 500$, we cannot claim that ICXP constructs more risk averse solutions than ICCsP for all reference random outcomes (for all values of the parameter ρ).

Figure 2 illustrates that our models obtain more risk averse optimal solutions when the level of risk aversion gets higher. Thus, the total cost is nonincreasing while the risk aversion parameters α and δ are increasing, i.e., while the level of risk aversion is getting smaller. Figure 2 also demonstrates the interaction between the local and global constraints. It is easy to see that for large enough δ values the global ICC constraint would be implied by ICCs (33) and it would become redundant. In this case increasing the value of δ would not affect the optimal solutions, since a certain level of service is ensured by the local constraints. Similarly, for large enough α values the global ICC constraint would be more significant and the optimal solutions are not affected by further increasing α values. Figure 2 is consistent with these observations; the total cost, the number of facilities and vehicles become constant after

Table 10 Results by ICXP for different reference random total unmet demand

Problem instances	For reference benchmark ρ	Total # of facilities	Ratio of total # of vehicles	Total cost	Total expected unmet demand
$n = 200$ $\|S\| = 500$	0.8	32	0.39	73,590.76	10.29
$n = 200$ $\|S\| = 500$	0.85	33	0.41	77,758.72	9.19
$n = 200$ $\|S\| = 500$	0.9	39	0.48	91,198.71	6.24
$n = 200$ $\|S\| = 500$	0.95	50	0.62	120,686.61	2.37
$n = 200$ $\|S\| = 3000$	0.85	32	0.34	82,632.97	8.43
$n = 200$ $\|S\| = 3000$	0.9	38	0.41	100,510.27	4.80
$n = 200$ $\|S\| = 3000$	0.95	53	0.55	142,386.77	1.61

some large enough δ and α values. In particular, these large enough values are $\delta = 0.052$ and $\alpha = 0.17$, respectively. These values can be justified by solving ICCsP after dropping the global ICC. For the optimal solution of the modified ICCsP, which involves only the local constraints with $\alpha = 0.2$ and $\alpha = 0.17$, we calculate the associated value of the global risk parameter δ for which the global constraint (34) is (externally) satisfied as 0.052 and 0.0399, respectively. That is why we observe that results stay constant around $\delta = 0.05$ and $\alpha = 0.17$ in Fig. 2.

Tables 7–10 provide results for several problem instances to show how the optimal solutions change with respect to different risk preferences. We also generate reference distributions to represent more risk averse preferences by increasing ρ, the proportion of demand that must be covered within $r_1 = 8$ minutes. It is not surprising that for larger values of ρ the model ICXP constructs more conservative solutions with higher total cost values.

7.7 Comparison with an existing model

We perform a computational study to compare our models to the one proposed by Beraldi et al. (2004). In general, Beraldi et al. (2004) consider the following probabilistic constraints:

$$P\left(\sum_{j\in N_i} x_{ij} \geq h_i(\omega),\ \forall i \in G(k)\right) \geq 1 - \epsilon_k, \quad k = 1, \ldots, K, \tag{65}$$

where K denotes the number of sub-area indexed by k, $G(k)$ denotes the components of the random demand vector associated with the group k and $1 - \epsilon_k$ is the prescribed probability level for the group k. They consider three cases described below as quoted from their paper: "(1) a global reliability system for the entire geographical territory (Jo); (2) an individual reliability system for each demand point (In); (3) an individual reliability system for sub-area (Jo-In)." Notice that for the first case, $K = 1$ and for the second case K is equal to the number of demand nodes (each group consists of an individual demand node). In the last case we have divided the demand points in (5) (Jo-In 5) or (10) (Jo-In 10) sub-areas.

As mentioned in Sect. 2, the reformulation presented in Beraldi et al. (2004) is not valid for a given set of scenarios. Therefore, in order to solve the problems which involve probabilistic constraints instead of our risk constraints, we need to consider another reformulation. A common approach introduces a binary variable ζ_s associated with each scenario $s \in S$. Then, for a group $k \in K$ we reformulate (65) as

$$\sum_{j \in N_i} x_{ij} \geq h_i^s (1 - \zeta_s), \quad \forall i \in G(k),\ s \in S, \tag{66}$$

$$\sum_{s \in S} p_s \zeta_s \leq \epsilon_k, \tag{67}$$

$$\zeta \in \{0, 1\}^{|S|}. \tag{68}$$

When the binary variable ζ_s takes value of 0, it is guaranteed by (66) that all the inequalities $\sum_{j \in N_i} x_{ij} \geq h_i^s,\ i \in G(k)$, hold. Constraints (66) and (67) require that demand at each node will be satisfied for a set of scenarios whose aggregate probability is at least equal to the enforced probability level $1 - \epsilon_k$. We derive a stronger formulation of (66)–(68) by employing a modeling approach proposed by Luedtke et al. (2010) to solve problems with joint probabilistic constraints. (For details please see Luedtke et al. 2010.) In Tables 11 and 12 we present results obtained for three cases considered by Beraldi et al. (2004) and also the results obtained by our models for several risk parameters. We also report the minimum of the resulting probabilities of satisfying local demands.

As seen from Tables 11 and 12 similar results can be obtained using probabilistic constraints or the proposed alternate risk measures by varying the associated risk parameters. Obviously, considering just individual probabilities leads to high unmet demand values. On the other hand, enforcing a single joint probabilistic constraint is too conservative. A rational decision maker would prefer an approach somewhere between these two extremes. Such a decision maker can enforce joint probabilistic constraints for sub-areas or can use ICCs or a stochastic dominance constraint depending on his/her risk preferences. As the performed computational study shows our proposed models and the associated solution methods enable the decision makers to solve relatively large problem instances.

Table 11 Comparative results for a problem instance where $n = 100$ and $|S| = 100$

Problems	Total # of facilities	Total # of vehicles	Total cost	Total expected unmet demand	Minimum of all local probabilities
Base problem	32	102	85,508.51	5.86	0.71
Worst case approach	72	229	222,779.59	0.00	1.00
Jo	63	207	195,487.23	0.22	0.99
Jo-In(5)	55	180	162,799.40	0.62	0.90
Jo-In(10)	49	161	142,576.17	1.12	0.90
In	33	106	92,416.44	4.35	0.90
ICCsP, $\alpha = 0.01$	61	199	181,955.12	0.37	0.97
ICCsP, $\alpha = 0.05$	45	148	130,488.77	1.55	0.95
ICCsP, $\alpha = 0.2$	41	137	116,091.63	2.28	0.92
ICXP, $\alpha = 0.01$	50	170	154,080.23	0.85	0.97
ICXP, $\alpha = 0.2$	45	147	127,929.99	2.01	0.89

Table 12 Comparative results for a problem instance where $n = 100$ and $|S| = 300$

Problems	Total # of facilities	Total # of vehicles	Total cost	Total expected unmet demand	Minimum of all local probabilities
Base problem	29	102	78,183.33	6.15	0.73
Worst case approach	*	268	*	*	*
Jo	76	242	223,909.80	0.13	0.98
Jo-In(5)	56	189	156,299.87	0.59	0.90
Jo-In(10)	48	160	127,193.54	1.16	0.90
In	32	111	84,612.15	3.78	0.91
ICCsP, $\alpha = 0.01$	59	195	165,053.65	0.50	0.98
ICCsP, $\alpha = 0.05$	45	149	118,316.89	1.46	0.94
ICCsP, $\alpha = 0.2$	39	114	103,143.52	2.15	0.93
ICXP, $\alpha = 0.01$	49	166	135,967.37	0.94	0.98
ICXP, $\alpha = 0.2$	45	152	120,072.41	1.65	0.92

*Not feasible; the total demand cannot be satisfied due to the upper bounds on the number of vehicles that can be allocated to each facility

8 Conclusion and future work

Risk measures should be incorporated into facility location and allocation problems in order to consider the inherent variability in the system and the decision makers' risk preferences. In this paper, we develop new stochastic programming models specifying alternate constraints on risk for the problem of designing an EMS system. We present a novel approach to EMS design problems by modeling risk through the integrated chance and stochastic dominance constraints. The decision makers can evaluate different location and allocation decisions with respect to the quality of service and costs by varying the risk parameters. The presented numerical results illustrate how the location and allocation solutions change with respect to the different risk preferences. One of the main disadvantages of the scenario approach is the requirement to limit the number of scenarios for computational reasons. We develop alternative formulations and a heuristic for the proposed single-stage models and by performing an extensive computational study we show that we can overcome the main drawback of the scenario approach and solve relatively large problem instances.

In practice, the total demand varies depending on the time of the day or the day of the week. Our models can be used to solve the problem of determining the facility locations and the number of vehicles allocated to each facility. Then, given the optimal number of facilities and the allocation of the emergency vehicles, modified versions of our models, where the objective function represents the total hourly cost, can be solved 168 times for each hour of a week. The optimal allocation decisions for each time period would provide us with the number of vehicles required at each facility for each hourly time period of a week according to the specified risk preferences. Then, existing methods for reallocating emergency vehicles and multistage crew scheduling may be applied to improve the EMS system design.

Part of our ongoing research is on developing methods to solve the two-stage versions of the proposed EMS design models for moderate size problem instances. The future research can focus on developing similar stochastic models which consider other features of an EMS system, such as different types of vehicles to serve different types of service requests. For example, EMS systems typically work with two types of service providers having different capabilities: basic life support units and advanced life support units. Finally, the dispatching rules that are related to the allocation of vehicles to specific nodes can also be modeled by using the proposed risk measures in the two-stage stochastic programming framework.

Acknowledgements The author would like to thank anonymous referees for their valuable comments and suggestions.

References

Ball, M. O., & Lin, F. L. (1993). A reliability model applied to emergency service vehicle location. *Operations Research*, *41*, 18–36.

Batta, R., Dolan, J. M., & Krishnamurthy, N. N. (1989). The maximal expected covering location problem: Revisited. *Transportation Science*, *23*, 277–287.

Beraldi, P., & Bruni, M. E. (2009). A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research*, *158*(1), 183–193.

Beraldi, P., Bruni, M. E., & Conforti, D. (2004). Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research*, *196*, 323–331.

Berman, O., & Krass, D. (2001). Facility location problems with stochastic demands and congestion. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: applications and theory* (pp. 331–373). New York: Springer.

Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research*, *147*, 451–463.

Daskin, M. S. (1983). A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, *17*(1), 48–70.

Daskin, M. S. (1995). *Network and discrete location: models, algorithms and applications*. New York: Wiley.

Dentcheva, D., & Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, *14*, 548–566.

Dentcheva, D., & Ruszczyński, A. (2006). Portfolio optimization with first order stochastic dominance constraints. *Journal of Banking and Finance*, *30*(2), 433–451.

Erkut, E., Ingolfsson, A., & Erdogan, G. (2008a). Ambulance location for maximum survival. *Naval Research Logistics Quarterly*, *55*(1), 42–58.

Erkut, E., Ingolfsson, A., & Budge, S. (2008b). Maximum availability/reliability models for selecting ambulance station and vehicle locations: a critique. Working paper. Available from http://www.business.ualberta.ca/aingolfsson/publications.htm.

Felder, S., & Brinkmann, H. (2002). Spatial allocation of emergency medical services: minimizing the death rate or providing equal access? *Regional Science and Urban Economics*, *32*(1), 27–45.

Fourer, R., Gay, D. M., & Kernighan, B. W. (1993). *AMPL: a modelling language for mathematical programming*. San Francisco: Scientific Press.

Gendreau, M., LaPorte, G., & Semet, F. (1997). Solving an ambulance location model by tabu search. *Location Science*, *5*(2), 75–88.

Goldberg, J. B. (2004). Operations research models for the deployment of emergency services vehicles. *EMS Management Journal*, *1*(1), 20–39.

Hadar, J., & Russell, W. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, *59*, 25–34.

ILOG (2006). ILOG AMPL CPLEX system version 10.0 user's guide. ILOG CPLEX Division.

Ingolfsson, A., Budge, S., & Erkut, E. (2008). Optimal ambulance location with random delays and travel times. *Health Care Management Science*, *11*(3), 262–274.

Klein Haneveld, W. K. (1986). *Duality in stochastic linear and dynamic programming*. *Lecture notes in economics and mathematical systems*, vol. *274*. New York: Springer.

Klein Haneveld, W. K., & Van Der Vlerk, M. H. (2006). Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science*, *3*(4), 245–269.

Lehmann, E. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, *26*, 399–419.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management Science*, *38*, 555–593.

Louveaux, F. V. (1986). Discrete stochastic location models. *Annals of Operation Research*, *6*, 23–34.

Louveaux, F. V. (1993). Stochastic location analysis. *Location Science*, *1*(2), 127–154.

Luedtke, J. (2008). New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, *19*(3), 1433–1450.

Luedtke, J., Ahmed, S., & Nemhauser, G. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming. Series A*, *122*(2), 247–272.
Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50–60.
Marianov, V., & ReVelle, C. (1995). Siting emergency services. In Z. Drezner (Ed.), *Facility location: a survey of applications and methods*. Berlin: Springer.
Marsh, M., & Schilling, D. (1994). Equity measurement in facility location analysis: a review and framework. *European Journal of Operational Research*, *74*(1), 1–17.
Müller, A., & Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. Chichester: Wiley.
Noyan, N., & Prékopa, A. (2006). A variant of the Hungarian inventory control model. *International Journal of Production Economics*, *103*, 784–797.
Owen, S. H., & Daskin, M. S. (1998). Strategic facility location: a review. *European Journal of Operational Research*, *111*(3), 423–447.
Prékopa, A. (1973). Contributions to the theory of stochastic programming. *Mathematical Programming*, *4*, 202–221.
Prékopa, A. (1995). *Stochastic programming*. Dordrecht: Kluwer Academic.
ReVelle, C., & Hogan, K. (1989). The maximum availability location problem. *Transportation Science*, *23*, 192–200.
Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE Transactions*, *38*(7), 537–554.