Tactical Robust Decision-Making Methodology: Effect of Disease Spread Model Fidelity on Option Awareness

Jennifer Mathieu[1], Mark Pfaff[2], Gary Klein[1], Jill Drury[1], Michael Geodecke[3], John James[1], Paula Mahoney[1], Georgiy Bobashev[3]

The MITRE Corporation[1]
jmathieu@mitre.org
gklein@mitre.org
jldrury@mitre.org
jhj@mitre.org
pmel@mitre.org

Indiana University – Indianapolis[2]
mpfaff@iupui.edu

RTI International[3]
mgeodecke@rti.org
bobashev@rti.org

ABSTRACT

We demonstrate a method of validating the utility of simpler, more agile models for supporting tactical robust decision making. The key is a focus on the decision space rather than the situation space in decision making under deep uncertainty. Whereas the situation space is characterized by facts about the operational environment, the decision space is characterized by a comparison of the options for action. To visualize the range of options available, we can use computer models to generate the distribution of plausible consequences for each decision option. If we can avoid needless detail in these models, we can save computational time and enable more tactical decision-making, which will in turn contribute to more efficient Information Technology systems. We show how simpler low fidelity, low precision models can be proved to be sufficient to support the decision maker. This is a pioneering application of exploratory modeling to address the human-computer integration requirements of tactical robust decision making.

Keywords

Decision-making, courses of action, modeling and simulation, agent-based model, equation model.

INTRODUCTION

In a world where emergencies can arise quickly, decisions frequently need to be made rapidly if they are to prevent additional damage and suffering. Such decisions must be tactical both in the generic sense (calculated, clever, and well planned) as well as in the military sense (directing immediate action). As recent natural disasters such as Hurricane Katrina have made abundantly clear, these decisions can be difficult to make against a backdrop of deep uncertainty (Lampert et al., 2003) due to having little current information under widespread infrastructure destruction.

Emergency response strategists such as pandemic influenza researchers have often used modeling and simulation as tools to forecast plausible futures, but there are two challenges to using such models for effectively supporting choices among available options in emergency decision making within today's especially volatile environment. The first challenge is dealing with a plan's sensitivity to input assumptions. Most plans require detailed inputs that involve a number of assumptions and predictions, which may or may not turn out to be true. When an emergency scenario is modeled, the indicated optimal course of action (COA) under one set of inputs and assumptions may be shown to be unsatisfactory under plausible changes to those values. For example, the pandemic influenza plans prior to H1N1 called for courses of action that only considered a high disease morbidity/mortality rate; as illustrated in the 2009 H1N1 pandemic, a lower morbidity rate made these plans sub-optimal. Indeed, elucidating such situations turns out to be a strength of using simulation modeling, not a weakness. The second challenge is maintaining agility by identifying simulation models that can both generate useful forecasts for decision making and execute in a viably short amount of time.

Reviewing Statement: This paper has been fully double blind peer reviewed.

Robust Decision Making (RDM, Lempert et al., 2003 and Chandrasekaran, 2005) addresses the challenge of sensitivity to input assumptions. RDM expands upon the concept of sensitivity analysis, using models to generate the plausible outcomes of making a particular decision under a broad range of different combinations of input values and assumptions. For a given course of action, RDM analysis evaluates the effects of the course of action, as expressed through variables in the model, in combination with all reasonable values of modeled variables that decision-makers *cannot* control. Even though finding, adapting, or developing appropriate models can be difficult, with such models, simulation runs for each potential course of action can generate a range of plausible futures. This "use of [a] series of such computational experiments to explore the implications of varying assumptions and hypotheses," has been called *exploratory modeling* by Bankes (1993).

Each plausible future can be scored in terms of the cost of enacting the decision, the costs of the consequences resulting from the decision, and any opportunity costs or additional costs that might occur in the future due to having taken the course of action. For example, if one was to provide vaccines to 75% of the population, the ideal cost for that COA could be computed as the total of the price of administering the vaccine doses, the cost of treating the people who become ill, the cost (defined by actuarial tables) of the resulting deaths, and the potential cost in illness and death from not having enough reserve vaccine for a potential future wave.

A set of simulation runs can be performed for each decision option using slightly different assumptions for the exogenous variables (those that are beyond decision makers' control) in combination with reasonable variation in the endogenous variables (determined by the course of action). The result of each run may have a different total cost. A *robust* option will generate relatively tightly clustered outcome costs across multiple simulation runs, which indicates that its outcomes are relatively impervious to perturbations in exogenous and endogenous variable values. In contrast, a *brittle* option will have forecasted costs that differ greatly across differing assumptions. In that case, the decision-maker can *drill down* into the more unfavorable outcomes to see if there is anything that can be done to mitigate the effects of the underlying conditions indicated by the model.

Using RDM with simulation modeling to forecast costs across a range of plausible futures, the decision maker can determine which decisions are more likely to stay within the range of acceptable costs, regardless of plausible variations in external factors not under the decision maker's control, and with reasonable variations in factors that are under his or her control.

While RDM has proven value as a strategic tool (Lempert, 2003), the current work is a pioneering effort to make RDM more practical as an agile tactical tool: enabling decision-makers to run RDM analyses fast enough to be of use in time-limited decisions. Clearly, exploratory modeling can involve many simulation modeling runs, which in turn can take a lot of time: on the order of days for some models. More complicated and precise models require more time to run, which can be an exponential function of the number and precision of the model variables. Consequently, we are developing a method for determining the most agile model with the lowest fidelity and precision necessary to produce results that would lead to the same decisions as would be made using the highest fidelity, most precise and slowest executing model.

While we have chosen the disease spread domain as a case study due to its critical importance and high penalty for incorrect decisions, we believe that this work is applicable to other domains such as electricity distribution or military command and control. Our efforts are aimed at developing a process so that others can run similar analyses for their own modeling work.

BACKGROUND AND RELATED LITERATURE

Using RDM normally involves depicting the results of the simulation models in a way that reveals the model's sensitivity to changes in input parameters and external conditions. The depictions are key to choosing a robust course of action. However, with a slight change in procedure, these depictions can also prove whether alternative model fidelity/precision levels are sufficient for tactical usage. Thus it is worthwhile providing more background on visualizing robust courses of action.

Using simulation models to generate plausible futures, then scoring and visualizing the results, allows us to translate probabilities into frequency distributions. Hoffrage and Gigerenzer (1998) showed that information presented as probabilities is harder to interpret than information presented as natural frequencies. Thus an RDM approach is likely to help both tactical decision-makers and those who are evaluating whether different models present decision options in the same order based on the distribution of their plausible scores.

Our emphasis on showing decision makers explicit representations of the available decision options was based in part by Hall et al.'s (2007) concept of the *decision space*. A decision space consists of information that shows decision-makers the relative desirability of the various available courses of action. Hall et al. contrast the decision space with the *situation space*: facts about the environment that constitute the information needed to attain situation awareness. To take action based on situation space information, decision makers need to determine what options are feasible and then relate those options to the situation space. In common situations, experts can immediately determine the correct option (Zsambok & Klein, 1997). However, in complex decision situations under deep uncertainty, inherently forming these situation awareness-to-option-awareness relationships becomes problematic. We have found that the situation space-decision space distinction is quite powerful: it has enabled us to focus more clearly on generating and presenting *decision ready* information rather than data that still requires additional mental processing before it can be used to inform decisions. We refer to this as providing *option awareness* (Drury et al., 2009).

A concrete example will help illustrate how decision space and situation space presentations differ. Consider the potential spread of H1N1 in Viet Nam as shown in charts and tables that describe the number of illnesses as a function of time and location (Boni et al., 2009). In this example, the information is quite detailed: numbers are broken down by duck and chicken owners; numbers of illnesses are shown on a daily basis; and a series of miniature maps show numbers of affected people on a province-by-province basis every twenty days. Yet all of this information describes the situation as it evolves: it is solely situation space information.

Other disease spread modeling studies do provide decision space information. For example, Dibble et al. (2007), Ferguson et al. (2006), and Germann et al. (2006) present similar information as Boni et al. (2009) but add charts and tables to show the numbers of illnesses assuming that specific courses of action are taken. Ferguson et al. (2006) describes the impact on the number of illness over time based on travel restrictions, several different vaccination strategies, and social distancing measures such as school closures and voluntary household quarantines. Germann et al. (2006) shows the impact of social distancing and travel restrictions, vaccination, and targeted antiviral prophylaxis as a function of time and also four different R_0 levels[1], an important exogenous variable. Information showing the effects of these different courses of action is firmly grounded in the decision space.

However, for tactical operational usage, the information presented to decision makers in these and other papers in the disease spread literature, is lacking in two critical ways. First, the studies do not provide a means of value scoring: that is, a ready means of determining which course of action is "best" in some externally-quantified manner. Studies often show bar charts or curves of numbers of people ill versus time; but the curve that represents the best course of action is not necessarily apparent. What is the basis for trade-off among options? Should the decision maker perform integrations to determine the curve that contains smallest total area – the smallest number of people ill? Or should they choose the curve that has the fewest high spikes, assuming that the spikes indicate periods of time when hospitals could become overloaded with peak numbers of sick people? Furthermore, should these comparisons be based purely upon visual inspection, or must the decision maker mentally perform complex calculations? When tuned to represent explicitly the parameters that are of importance to decision makers, the results of a value scoring model can lessen cognitive burden by providing visualizations that inherently discriminate the relative robustness of the options.

Second, the studies previously cited do not provide means of interactively drilling down into information aggregations to view constituent data items and additional detail. Drilling down, for example, can reveal to the decision maker the exact number of illness cases represented by a colored dot or area on a map such as was presented in Boni et al. (2009) and Germann et al. (2006).

In contrast, the RDM approach uses a value scoring model to provide a means of comparing one COA to another. Moreover, extending this method, the scoring results coupled with the visualization enables analysts to compare the results of *different* models by revealing whether those models result in different orderings of options based on their forecasted value scores. If those orderings are the same, then decision makers would be likely to make the same decision using any of the models. If so, then they can validly use a computationally more agile model of lower fidelity or precision to support decision making in rapid response, tactical situations.

Levels at which the decisions are made often determine the granularity and complexity of a model. Naturally the granularity unit should be smaller than a decision unit. For example, is a decision is to be made about a particular town or village a model with smallest units representing countries will be of little use. However increased granularity also increases complexity and the number of parameters to consider. The difference in model complexity is reflected in the computational resources to run the simulations. Growing computation power opens horizons for the development of more complex systems; however there is a need to understand whether the decision making actually benefits from such increased complexity.

---

[1] R_0 is "the average number of secondary infections caused by a single typical infected individual among a completely susceptible population" (Germann et al., 2009, p. 5935).

TACTICAL ROBUST DECISION MAKING METHODOLOGY

An eight-step exploratory modeling process can be used to determine an acceptable model fidelity level for generating the decision space needed for tactical robust decision making in a given domain:

1. Define options or courses of action among which decision makers are trying to distinguish;
2. Define endogenous and exogenous variables relative to the courses of action;
3. Develop or adapt forecasting models that execute at different levels of detail or fidelity;
4. Define a value scoring function for evaluating model forecasts;
5. Define how to manipulate model precision;
6. Design and run an exploratory modeling experiment to generate a landscape of plausible futures;
7. Visualize the scored landscapes of plausible futures – enabled by interactive tools; and
8. Analyze the results statistically.

Step 1: Define options or courses of action

Procedure: Describing the decision space for a particular domain helps to drive option awareness. Defining options or courses of action that decision-makers are trying to distinguish between is the first step in tactical robust decision making. Example courses of action are decisions on: amount of resources to acquire, allocation of resources, distribution of resources, etc.

Example: For contagious disease spread models, options include: (1) recommending social distancing — a voluntary behavior in which people stay home to avoid propagating the disease, (2) amount of medication (e.g., antivirals) to obtain and store for treatment of the illness, (3) amount of vaccines to obtain to prevent further illness, and (4) distribution strategy for the vaccines. If there are four courses of action each at two levels then there are 16 possible combinations.

Step 2: Define endogenous and exogenous variables

Procedure: Both endogenous and exogenous variables relative to the course of action need to be defined at various levels. Each level has a certain amount of assumed variability or uncertainty.

Example: For the courses of action identified above, voluntary social distancing and vaccination strategy can be defined as yes/no (they are either used or not used) and vaccination can occur at primary provider location or via an en mass clinic. No variation was used since the choices were defined as binary. The amount of medication and vaccinations available were selected to be low and high at 10% and 50% of the population for antivirals and 25% and 75% of the population for vaccinations. Five exogenous variables were defined: initial population susceptible to disease (95%), disease infectivity (high), mortality rate (1%), percent of population that seek medical attention (30%), and percent of population that complies with recommend social distancing (70%). For this experiment, the exogenous variables were evaluated at only one level. The uncertainty distributions defined for the endogenous and exogenous variables were either truncated normal or triangular distributions, with a range of ±10% of the variable value since data was not available on the actual uncertainty for the variable.

Step 3: Develop models of different fidelities

Procedure: To determine the lowest fidelity model to support decision making for a given domain, models of various fidelities must exist or be developed. Simulation models can be constructed at different levels of detail to describe the same phenomena. Often, different modeling approaches (e.g., agent-based modeling, discrete event simulation, system dynamics, etc.) are used for different levels of fidelity. However, one can use the same modeling approach at two levels of fidelity by using more detail in one model than the other. Factors that can be manipulated to increase or decrease fidelity include complexity of the governing equations, number of agents, frequency of updating model calculated variables, etc.

Example: The high fidelity model adapted for this study is a hybrid disease spread (Beeker, 2009) and process (Mathieu et al., 2009) model developed by [Company 1] and the lower fidelity disease spread model adapted for this study was developed by [Company 2] (Epstein et al., 2007). The high fidelity model is a hybrid agent-based, discrete-event model (referred to hereafter as the agent-based model) at the person level, and the lower fidelity model is an equation-based model (which we will refer to as an equation model) at the population level. The courses of action and parameters were based on Epstein et al. (2008); however, the model was adapted for a hospital catchment area. Note that only courses of action that could be tested using the equation model were evaluated so that we could directly compare the results of the two types of models.

Step 4: Develop value scoring function

Procedure: Define measures of effectiveness and other factors that contribute to a value scoring function that will be used to evaluate model forecasts. The models drive the distribution of costs for each course of action in the decision space. To map this situational information into the decision space a way of aggregating the results of each option is needed. Cost is a common measure: it is nearly always possible to assign costs to the activities involved in a course of action and the results of that action.

Example: The cost of purchasing antivirals, purchasing vaccines, and human productivity loss were used to define the value scoring function. Productivity loss was calculated for both people-sick days and death, with a very high cost assigned to death.

Step 5: Define model precision

Procedure: Precision is defined by the number of agents (only for an agent-based model), number of simulation replicates (low vs. high number), and variability in sampling the distribution defined for given endogenous and exogenous variables. In the simplest case, two models of varying level of fidelity can be used, each with two levels of precision.

Example: For the agent-based model, the high precision case had 800 families (~2170 agents) with 30 replicates, and the low precision case had 400 families (~1085 agents) with 15 replicates. Analogously the equation model high precision case had a sample rate every 0.5 days with 30 replicates, and the low precision case had a sample rate every one day with 15 replicates. The high precision cases utilized the selection from the normal and triangular distributions and the low precision cases limited the selection to three values from the normal and triangular distributions.

Step 6: Create an experimental design to generate a landscape of plausible futures

Procedure: This is the core of this exploratory modeling process. In this step, an experimental design is developed that will assess the models' sensitivity to endogenous and exogenous variables and values that are critical to option awareness. The particular variables and values chosen must be based on knowledge of the application domain. As can be seen from the example below in Table 1, a complete factorial design for even a relatively small number of variables can require a large number of model executions. Other designs (e.g., Latin Square) can reduce the required combinations, but with a loss of statistical sensitivity. So, there must be a tradeoff evaluated among the number of critical variables, experimental design and statistical sensitivity. We recommend making a few dry runs to determine the processing time needed for an individual execution of the model, so that the final design can be informed by the practical need to finish the runs in an acceptable amount of time.

Given a design, for each value or values selected, a range and probability distribution needs to be defined. A sampling-based sensitivity analysis is then executed for each of the models at each level of precision. Each model is run multiple times, each time with new combinations of values sampled from the specified ranges, with specified probabilities. Each simulation run results in the generation of one plausible future, with multiple executions resulting in a landscape of plausible futures. Each of these futures is scored according to the value scoring function developed in step 4.

Example: As shown in Table 1, this study has four models (agent-based model (ABM) High Precision, ABM Low Precision, equation model (EM) High Precision, EM Low Precision) with each of the four courses of action (Social Distancing, Percent Antivirals, Percent Vaccinated, Vaccination Strategy) tested at two levels. The high precision models have 30 replicates resulting in 480 simulation runs to achieve all combinations, and the low precision models have 15 replicates resulting in 240 simulation runs to achieve all combinations. In addition, each simulation run includes a sampling from the critical exogenous variables value ranges as well.

Step 7: Visualize Results

Procedure: Once the COAs and their costs have been generated for the different models, the next step is to examine the results to see whether the different models lead to differences in the relative desirability of the COAs. In other words, determine whether decision-makers would be led to a different choice depending on the model used. This effort starts with a visual inspection and then moves to a statistical analysis in step 8. Visualizing COAs is also useful to decision makers (Drury et al., 2009), so this step has utility beyond helping to determine a reasonable model to use for creating the decision space.

Because the relative costs of COAs should be easily understood, the manner in which COAs are depicted will depend upon the expectations and backgrounds of the people who will be using those visualizations. Specifically, the visualizations should be clear depictions of the shape of the frequency distributions, highlighting their central tendencies and variability. If differences in the models impact key elements of the frequency distribution, these changes may affect how people make decisions and their confidence in those decisions. Thus it is important to examine the results generated by the various models in a way that makes any differences clear.

Model	Families / Time Step	# of Replicates	Social Distancing	Percent Antivirals[4]	Percent Vaccinated[4]	Vaccination Strategy	# of Runs[5]
ABM[2] High Precision	800 families	30	no yes	10 50	25 75	daily one-time	480
ABM Low Precision	400 families	15	no yes	10 50	25 75	daily one-time	240
EM[3] High Precision	0.5 day	30	no yes	10 50	25 75	daily one-time	480
EM Low Precision	1 day	15	no yes	10 50	25 75	daily one-time	240

Table 1. Experimental Design[1]

Table notes:

1. For all models, percent susceptible = 95, percent who report illness = 30, mortality rate = 1, infectivity = high, and percent who comply with social distancing (when used) = 70.
2. ABM = Agent Based Model
3. EM = Equation Model
4. The uncertainty distributions defined for the endogenous variables relative to the courses of action were truncated normal distributions, with a range of ±10% of the variable value since data was not available on the actual uncertainty for the variable. The high precision cases utilized the selection from the normal distribution and the low precision cases limited the selection to 3 values from the normal distribution.
5. Number of runs is calculated by the number of replicates x 2 (the number of social distancing levels tested) x 2 (the number of percent antivirals levels tested) x 2 (the number of percent vaccinated levels tested) x 2 (the number of vaccination strategy levels tested).

Example: We developed a prototype data exploration tool that displays data in the form of Tukey's (1977) statistical box-plot visualizations. Showing a set of box-plots to represent decision options is consistent with the "small multiples" visualization technique (Tufte, 1990), which is comprised of a series of simple graphics that differ in only a few key ways. While there are more sophisticated visualization techniques, box-plots are sufficient for our case because they can enable analysts and decision makers to compare five characteristics of the data distributions and they are readily understood by the intended user population (Drury et al., 2009).

Each box-plot shows the median cost for one COA as a line within the box, and the box itself encloses the results ranging from the 25th percentile to the 75th percentile for that COA (the range enclosed by the box is known as the Inter-Quartile Range, or IQR). The "whiskers" each delineate the runs that are more than 1.5 times the IQR range above and below the box (in Figure 1, these are the two horizontal lines outside the boxes). Beyond the whiskers the remaining runs, if any, are known as outliers and they are denoted by dots in Figure 1.

Note that this display approach does not explicitly show the effect of time on the number of illnesses. The value scoring model adds the cost of the resources expended executing the COA to the direct and indirect economic costs of the illnesses and deaths[2] that occur over the entire timeline of the disease spread incident. Thus, all

[2] Note that monetized costs of death are based on actuarial data such as that used by insurance companies.

impacts of a COA over time are displayed as one box-plot that is an aggregation of all plausible results over the entire timeframe.

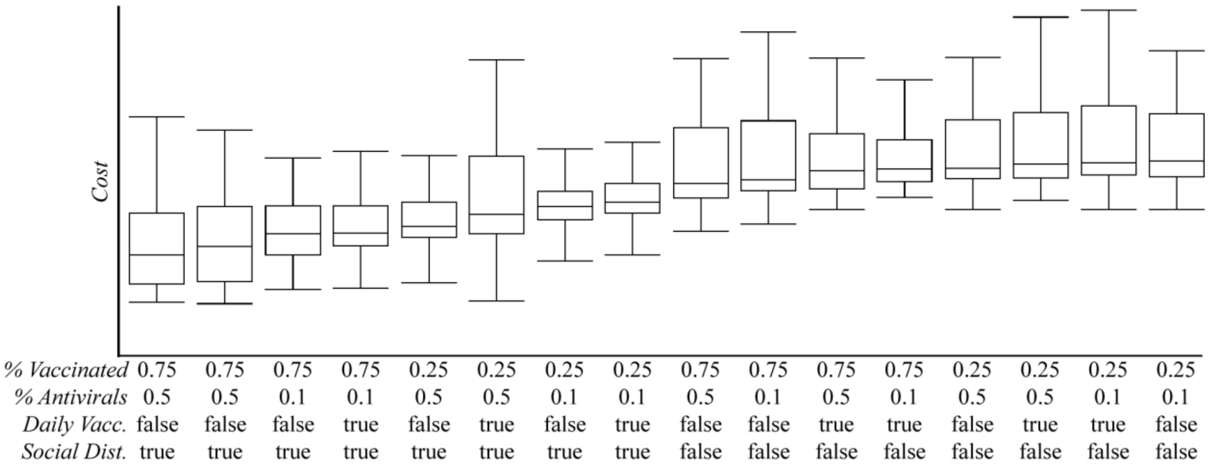

Figure 1. Box-plots for 16 combinations of the four courses of action sorted by median cost

The visualizations can help analysts and decision-makers to explore alternatives in several ways. First, they can see which box-plots are low and tight, indicating that the decision choices they represent would be low cost in virtually all circumstances. We say that this level of understanding corresponds to option awareness level 1. Second, drilling down into the box-plots reveals whether particular circumstances lead to costs that are higher or lower than expected. For example, if one case in a particular course of action involves vaccine distribution problems that result in unacceptably high costs, the decision-maker may still choose this course of action but take steps to ensure that delivery problems do not occur. Because it results in an understanding of the underlying data, examining the data at the drill-down level results in option awareness level 2.

Step 8. Analyze statistically with the aid of the interactive exploration tool

Procedure: A visualization method such as that described above facilitates ranking the COAs based on their value-score distributions. A statistical test such as Kendall's W can be used to determine concordance among the models' ranked COAs. Kendall's W addresses the trend of agreement, i.e. whether, on average, COAs appear in each model's rankings in more or less the same order.

If concordance is strong, then one can operationally use the lowest fidelity and precision model because its use will not result in substantially different decisions from those made using the highest fidelity and precision model. If concordance is not strong, then one can eliminate the model(s) that do not substantially agree with the highest fidelity/precision model; and operationally use the lowest fidelity/precision model of the remaining group. Note that agreement is most important among a few best options: decision makers will rarely choose the 14^{th} ranked option out of 16, for example.

Example: In the case of the disease spread models, the medians, upper quartiles, lower quartiles, maximums, and minimums for all box-plots were ranked from lowest to highest cost. A "best three out of five" rule was applied to rank the options according to these five features. Ties were settled by whichever option had the lowest median. This process resulted in a list of options for each model ranked in a manner akin to a user's visual analysis of the box-plots, as described in step 7, above.

Overall concordance in the dataset described in step 6 was quite strong, indicating that all four models agreed strongly on which options were good and which were bad, Kendall's $W = 0.96$, $p < .001$. Kendall's W is a non-parametric measure of inter-rater reliability which ranges from 0, meaning completely random unrelated responses, to 1 which indicates perfect agreement among raters. Note that the ranking trends were consistent among all models even though a Wilcoxon Rank-Sum test showed a main effect for the type of model (Wilcoxon W $(n_1 = n_2 = 720) = 378573$, $p < .001$), with the agent-based models showing a higher overall cost[3]

[3] Cost is reported in millions of dollars (US $)

(*M* = 1011.76) than the equation models (*M* = 730.78). Agreement was strongest with the highest-ranked options, although the only one all four models agreed upon was the very worst option, which used the lowest levels of vaccination and antivirals, administered daily vaccines, but did not employ social distancing. After removing that point as an outlier, there was a strong correlation between rankings and consensus between models, $r(13) = .54$, $p < .05$. The top four rankings are shown in Table 2.

Pairwise comparisons indicated how similar the models' ranked lists were to each other (i.e. did they agree on #1? did they agree on #2? and so on for all 16 rankings). This was measured with Kendall's τ, a nonparametric correlation measure ranging from -1 (opposite rankings) to 1 (identical rankings). This metric gets more directly at the ordered presentation of the results and to what degree each model displayed a significantly different order of options. Similarity in rankings was strongest within model types, second strongest for comparable levels of precision, and not significantly similar for models differing in both type and precision (see Table 3).

Non-parametric correlations were run to examine variations in the relationships between individual model factors and overall cost for each model (see Table 4). All models agreed on the strongest factors (social distancing and the percent of the population vaccinated). The vaccination strategy, however, was unanimously the least significant. However, artifactual differences between the two types of models prevent direct comparison of some underlying factors.

	Model			
Rank	ABM – High Precision	ABM – Low Precision	EM – High Precision	EM – Low Precision
1	(12) – 0.75 0.1 true true	(14) – 0.75 0.5 false true	(14) – 0.75 0.5 false true	(16) – 0.75 0.5 true true
2	(16) – 0.75 0.5 true true	(16) – 0.75 0.5 true true	(16) – 0.75 0.5 true true	(14) – 0.75 0.5 false true
3	(14) – 0.75 0.5 false true	(10) – 0.75 0.1 false true	(10) – 0.75 0.1 false true	(10) – 0.75 0.1 false true
4	(06) – 0.25 0.5 false true	(12) – 0.75 0.1 true true	(12) – 0.75 0.1 true true	(12) – 0.75 0.1 true true
LEGEND: (Option Number) – %Vaccination (0.25/0.75), %Antivirals (0.1/0.5), Daily Vaccine (true/false), Social Distancing (true/false)				

Table 2: Comparison of the top four option rankings for the four models

Model	ABM – High Precision	ABM – Low Precision	EM – High Precision	EM – Low Precision
ABM – High Precision	-			
ABM – Low Precision	.59**	-		
EM – High Precision	.47*	.35	-	
EM – Low Precision	.35	.43*	.55**	-
* $p < .05$; ** $p < .01$				

Table 3: Kendall's τ for pairwise comparison of option rankings for the four models

	Courses of Action			
Model	Social Distancing	Percent Antivirals	Percent Vaccinated	Vaccination Strategy
ABM – High Precision	-.65**	.02	-.16**	.04
ABM – Low Precision	-.54**	-.05	-.19**	.03
EM – High Precision	-.84**	-.12*	-.30**	.11*
EM – Low Precision	-.82**	-.24**	-.26**	.11
* $p < .05$; ** $p < .001$				

Table 4: Spearman's r for regressions of COA factors against the overall cost

DISCUSSION

As can be seen in Figure 1, the exploratory modeling approach of RDM not only allows us to contrast four different models, but allows us to view a combined decision space generated by all of the models. The RDM analysis of this combined space shows that the most robust option for preventing the spread of a contagious disease is social distancing. The box plot visualization of this decision space clearly illustrates that social distancing in combination with any other option results in not only a lower median cost, but also a lower maximum cost, a lower minimum cost and a lower inter-quartile range.

Contrasting the individual models, the individual decision spaces of all four models very strongly agree on the ordering of the options based on robustness. The inter-rater concordance, indicated by Kendall's *W* of 0.96, shows a remarkably strong level of agreement between models, If there was significant disagreement at this level, it could indicate systemic inconsistencies in one or more of the models.

When we evaluate the similarity of the options' rankings in more detail, we do find some discrepancies in the order. As shown in Table 3, a highly significant level of correlation exists between the two agent-based models and between the two equation models across precision levels. Interestingly, across the different model-types, there are statistically significant correlations between models of the same level of precision. This suggests that one could satisfactorily support the decision space using the more agile lower precision models. This suggestion is further supported when we examine the individual options.

At the individual option level, Table 2 shows that all of the top three options for each model include social distancing, and a 75% level of vaccination; two out the three top options for each model includes 50% anti-viral distribution level. So if a decision maker used the more agile low precision equation model (requiring 8 hours to run 480 simulations) and chose the most robust option (#16) identified by that model, this is the second most robust option indentified under the higher fidelity, high precision, and slower agent-based model (requiring 2 days to run 480 simulations). Even under the agent-based model, Option 16 is forecast to have only slightly higher median cost (about 4%) than Option 12, which is the one the agent-based model indicates as most robust. In addition, choosing Option 16 instead of Option 12 would not lead to a significant change in the course of action – it would simply increase the distribution of antivirals, an action that is recommended by two-thirds of the most robust options from all of the models. The exploratory modeling process verified that the low precision equation model would generate a satisfactory decision space for choosing among available options. It is significant to find that agreement is highest among the top-ranked options, which arguably would be the only ones considered by an actual decision maker. Disagreement about the more undesirable options is less likely to have an impact in practice.

When we look at the relationship between underlying factors and costs, we also find only minor differences among the models. Table 4 indicates that all four models agree on the two most significant factors: social distancing and percent vaccinated. We see that in general the agent-based model results in lower correlations between each factor and costs. Examination of the model results shows that this appears to be due to greater variability in the landscape of outcomes generated by the agent-based model. There is no case where models differ in the direction of a significant correlation. The results also show that in each model, the endogenous factors are mitigating the impact of the disease. This is indicated by the relatively low (albeit statistically significant in three models) correlation between the range of mortality rates of the disease and cost. Therefore, if one drills down to examine underlying causes of forecasted outcomes, one finds essentially the same relationships in all models. These results verify that even at the relationship level, the low precision equation model would yield satisfactory support for identifying the mitigating strategies tested.

CONCLUSIONS

These results illustrate the utility of this eight-step exploratory method for determining the most agile model. In this case the method verified that two very differently constructed models, at two different levels of precision, are equally satisfactory at generating a useful decision space for options that can be represented in all models. For assessing such options, the more agile equation model is a viable choice.

Four elements were combined for the first time in this research. The first element was the use of *exploratory modeling and sensitivity analysis* to generate landscapes of plausible futures. This was combined with applying a *frequentist* approach to depicting the value scoring of those futures in a *decision space*. Finally, the resulting decision spaces could be shown to be provably equal by applying *statistical analyses of the concordance* of the models' ordering of options.

While developing useful forecasting models to support a robust decision-making process can be challenging, our previous work (Drury et al., 2009) has shown RDM improves decision making and decision confidence. The current research shows that by evaluating underlying models in terms of option awareness, needless model fidelity and precision can be eliminated. Consequently, one can reduce development, maintenance, and execution costs, which will in turn contribute to more efficient Information Technology systems. This new application of exploratory modeling will provide decision makers with provably valid agile models that will make tactical robust decision making practicable.

ACKNOWLEDGMENTS

This work was supported in part by The MITRE Corporation projects: 19MSR062-AA, 45MSR019-AA, 45MSR026-FA, 1908M100-AA, 05AAV010-J5, and 45MSR001-A9. We would like to thank MITRE staff members Meredith Keybl and Jeanne Fandozzi for supporting the modeling effort, and Nathan Rackliffe (MITRE) and Loretta More (Pennsylvania State University) for work on interactive box-plot visualizations.

REFERENCES

1. Bankes, S. (1993). Exploratory modeling for policy analysis, *Operations Research*, 41(3), 435-449.
2. Beeker, E. (2009). The MITRE epidemiological model. *System Engineering Process Office (SEPO) Collaborations*, 6(1), 9, available at www.mitre.org/work/sepo/pdfs/collaborations_vol6_num1.pdf.
3. Boni, M. J., Manh, B. H., Thai, P. Q., Farrar, J., Hien, T. T., Hien, N. T., Kinh, N. V. and Horby, P. (2009). Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses (2009). *BMC Medicine*, 7(3), doi 10.1186/1741-7015-7-43.
4. Chandrasekaran, B. (2005). From optimal to robust COAs: Challenges in providing integrated decision support for simulation-based COA planning. Laboratory for AI Research, The Ohio State University.
5. Dibble, C., Wendel, S., and Carle, K. (2007). Simulating pandemic influenza risks of US cities. In *Proceedings of the 2007 Winter Simulation Conference*, Washington, D.C.
6. Drury, J. L., Klein, G. L., Pfaff, M., and More, L. (2009). Dynamic decision support for emergency responders. *Proc. of the 2009 IEEE Technologies for Homeland Security Conference*, Waltham, MA.
7. Epstein, J.M., Goedecke, D.M., Yu, F., Morris, R.J., Wagener, D.K., and Bobashev, G.V. (2007). Controlling pandemic flu: The value of international air travel restrictions. *Public Library of Science One*, 5(e401).
8. Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(27), doi 10.1038/nature04795.
9. Germann, T. C., Kadau, K., Longini, Jr., I. M., and Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proc. of the National Academies of Sciences*, 103(15), 5935 – 5940, 11 April 2006.
10. Hall, D. L., Hellar, B. and McNeese, M. (2007). Rethinking the data overload problem: Closing the gap between situation assessment and decision making, *Proc. of the 2007 National Symposium on Sensor and Data Fusion (NSSDF) Military Sensing Symposia (MSS)*, McLean, VA.
11. Hoffrage, U. and Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, 73(5), 538 – 540.
12. Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. RAND MR-1626. Santa Monica, Calif.: The RAND Corporation.
13. Mathieu, J., James, J., Mahoney, P., Fandozzi, J. (2009). Using agent-based approaches to improve disaster response preparedness. *System Engineering Process Office (SEPO) Collaborations*, 6(1), 12. Available at www.mitre.org/work/sepo/pdfs/collaborations_vol6_num1.pdf.
14. Tufte, E. R. *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.
15. Tukey, J. W. *Exploratory Data Analysis.* Reading, Mass: Addison-Wesley, 1977.
16. Zsambok, C.E. & Klein, G. (1997). *Naturalistic Decision Making*. Lawrence Erlbaum Assoc., Mahwah, New Jersey.