2010-03

# Team 8: Using Data Farming Techniques for Health Care Policy Analysis

Reitter, Norman; Widdis, Dan; Lai, Kah Wah

https://hdl.handle.net/10945/35685

# Team 8: Using Data Farming Techniques for Health Care Policy Analysis


## TEAM 8 MEMBERS

Norman Reitter
Dan Widdis
*Concurrent Technologies Corporation (CTC), US*
Kah Wah Lai
*Singapore*


## INTRODUCTION

Health care analysis has traditionally focused on understanding the impact of single intervention programs on single risk factors. Extensive research has been done on individual social risk factors that lead to disease. However, risk factors do not act independently. New research is required to understand the inter-relationships between environmental influences, social influences and human decisions across many risk factors. In addition, requirements are emerging to use a systems approach to analyze multi-factor intervention policies and the combined impact on overall population health and medical costs. This historical research approach and emerging needs have set up an environment that is ripe for using data farming techniques – large scale, efficient experimental design; creating data using modeling and simulation techniques; and a variety of statistical modeling methods to understand results.

In this paper, we discuss CTC's focus on using advanced analytical techniques for health care policy analysis. Specifically, we focus on progress made during the IDFW 20 in verifying an agent based simulation (ABS) model developed in a NetLogo® software program and refining a data farming approach for using the model for analysis.

### Background

Over the past year, CTC, a non-profit scientific applied research and development corporation, began to develop an approach to use agent based simulation as part of a data farming approach to provide enhanced research for health care policy analysis. As part of its overall research effort, CTC has focused on a holistic solutions approach to providing systems analysis including:

- Research health care policy areas and integrate specific focus area data into formats usable for continued research
- Develop ontology models to represent the interrelationships between the human decision environment and influencers on human decisions (e.g. social networks, intervention programs, etc.)
- Develop an agent based simulation (ABS) model to represent interactions and assess future impacts of intervention policies on population disease rates
- Apply data farming techniques to the combined solution approach to analyze policies to support trade-off decisions.

### Team Objectives

The IDFW 20 provided an opportunity for CTC analysts to verify the ABS model using the data farming techniques and leveraging SEED Center for Data Farming and IDFW 20 participant expertise to verify this approach and the ABS model. Pre-workshop objectives included:

- Develop an efficient experimental design to evaluate multiple health risk factors and understand their impacts on population health
- Use an agent based simulation model to harvest data for exploration and identification of potential intervention opportunities
- Evaluate the effects of single vs. multi-factor intervention policies on population health

### Problem

CTC's research focused on answering the question "How do intervention policies impact population level characteristics?" The team focused on analyzing individual smoking characteristics, the impact of social networks to influence decisions to start or stop smoking, and the effectiveness of smoking intervention programs on reducing the overall population smoking rates.

### Parameters and Measures of Effectiveness

For this part of our research, we use the percentage of smokers in the population as our main effect and six types of intervention programs:

- **ASPIRE** - Computerized smoking prevention curriculum: school-based self-study program
- **ESFA** - European Smoking prevention Framework Approach: integrated classroom with teacher, advertising, journalism
- **ASSIST** - A Stop Smoking in Schools Trial - school based, peer-led
- **PPBI** - Pediatric Practice-Based intervention - healthcare provider and peer-based
- **National Truth Campaign** - Advertising campaign and youth advocacy
- **SCYP** - Smoking Cessation for Youth Project

We vary intervention program coverage (or influence) over the population from 0% to 100%.

## APPROACH

During IDFW 20, CTC's approach was to maximize use of data farming experts at the SEED Center and leverage expertise from other IDFW participants. To support this activity, prior to the workshop, the CTC team developed an ABS model using NetLogo® and collected data to support the model including population demographics and impacts of intervention programs on reducing smoking rates. We also calculated a variety of odds ratios for use in the ABS model and modeled the influence of an individual's social network on their chance of becoming a smoker or ceasing smoking.

The team's activities during the workshop, in figure 1, show how the ABS model evolved and a robust DOE developed.


**Figure 1. Team 8 Activities during IDFW 20**

As a result of our pre-work, we were able to run the ABS model using an initial experimental design on Monday. From this baseline, we were able to improve the model and the experimental design and make subsequent model runs. On Tuesday we enhanced the DOE to more fully explore the sample space. This activity allowed us to make another set of simulation model runs and analyze results starting on Wednesday. From this point, we were able to explore specific parts of the model in more detail to increase our understanding of the results.

### ABS Model in NetLogo®

The CTC Team used the NetLogo®[1] software modeling language to develop an ABS model to support this project. Figure 2 shows a snapshot of the model version

In this model, each agent represented an individual. The individual agent had the following characteristics:

- Age
- Gender
- Race
- Relationships with other agents
- Smoker status (never, former, current)

Each agent maintained its smoker status at the age of 30 years for the remainder of its life. Life expectancy for each agent is based on actuary tables and current smoking status resulting in a chance that the agent dies each year based on current age and status.

The model uses a state-based probability of changes based on a set of odds ratios, developed from a significant amount of research from open source health research journals, to determine when an agent changes from one smoking state to another. The agent's social network, or set of peers, influences whether or not an agent changes smoking state.

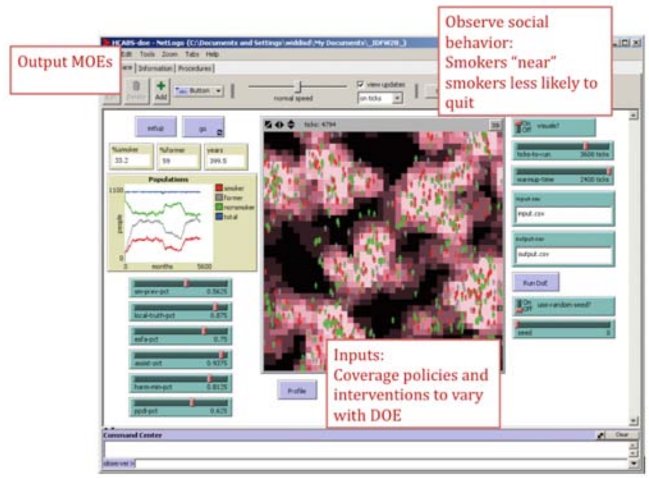



**Figure 2. Picture Snapshot of ABS model in NetLogo® at beginning of workshop**

The simulated environment lasts for approximately 300 years. We started the workshop considering 250 agents and expanded to 1000 agents by the end of the workshop.

Intervention programs were applied for a specified period of time over a portion of the population (from 0% to 100%), influencing the odds that an individual agent would change smoker states.

### DOE Development

Our DOE evolved over the course of the week, resulting in a denser, robust examination of our desired sample space. Figure 3 shows how each iteration of our experimental design improved coverage and density of sample points within the design space.

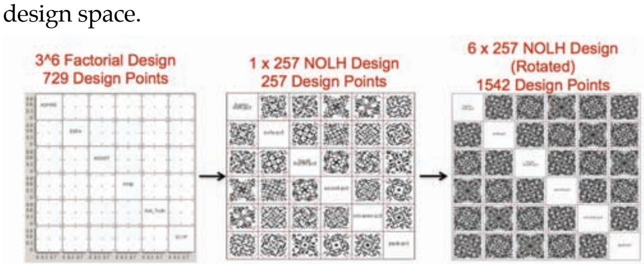


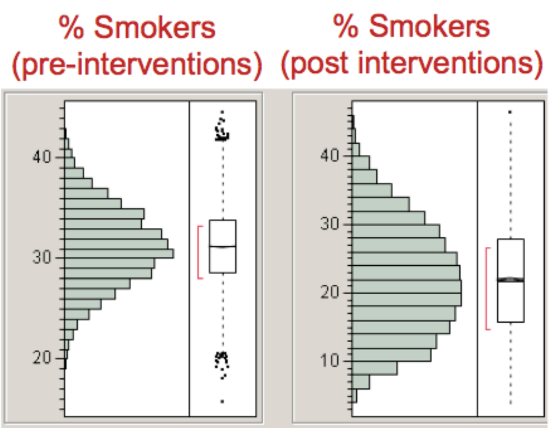

**Figure 3. Evolution of Experimental Design**

[1] http://ccl.northwestern.edu/NetLogo/

Our initial design used a full factorial considering each of the six interventions at three levels (0, .5, 1) resulting in 729 design points, or simulation model runs. We improved our design by using a Nearly Orthogonal Latin Hypercube (NOLH) model based on the SEED Center NOLH spreadsheet.[2] This provided a more robust sampling of our design space, with a reduction of 257 design points, however continued to expose some gaps based on the resulting combinations of intervention coverage inputs. We were also concerned that results showed possible dependencies between interventions-based sampling patterns. After conferring with SEED Center staff on how to more completely fill our design space, we used a rotated NOLH design, resulting in 1542 design points (simulation runs) and much richer sampling space represented by the right hand side design in figure 3. This final DOE allowed us a robust and efficient sampling plan to examine all combinations of intervention programs at many different levels (ranging from 0…1).

## RESULTS

Throughout the workshop we compared the percentage of smokers within a population prior to applying the interventions and then once the population smoking percentage reached a steady state after the interventions were applied over a period of time.




Quantiles (pre-interventions)

| | | |
|---|---|---|
| 100.0% | maximum | 44.800 |
| 99.5% | | 41.201 |
| 97.5% | | 38.843 |
| 90.0% | | 36.170 |
| 75.0% | quartile | 33.871 |
| 50.0% | median | 31.174 |
| 25.0% | quartile | 28.571 |
| 10.0% | | 26.210 |
| 2.5% | | 23.770 |
| 0.5% | | 21.338 |
| 0.0% | minimum | 15.789 |

Quantiles (post interventions)

| | | |
|---|---|---|
| 100.0% | maximum | 46.748 |
| 99.5% | | 41.559 |
| 97.5% | | 37.600 |
| 90.0% | | 32.773 |
| 75.0% | quartile | 27.823 |
| 50.0% | median | 21.656 |
| 25.0% | quartile | 15.702 |
| 10.0% | | 11.489 |
| 2.5% | | 8.434 |
| 0.5% | | 6.024 |
| 0.0% | minimum | 3.252 |

**Figure 4. Example Model Output: % Smokers Pre and Post Interventions**

Figure 4 shows an example of the distribution of percentage of smokers before and after the interventions were applied. In this example, the smoking population shifted ten percentage points from 31% to 21% of the total population.

With the effectiveness of each intervention being fundamental to our research, we next evaluated how each intervention acted independently on the reduction in the percentage of smokers in a population. In figure 5 we show an example of how one of the interventions, in this case the ASSIST intervention program, dominates the other interventions shown by a significantly steeper positively increasing slope when varied over increasing levels of coverage across the population (from 0% to 100%) with the most impressive influence when applied to over 50% of the population.

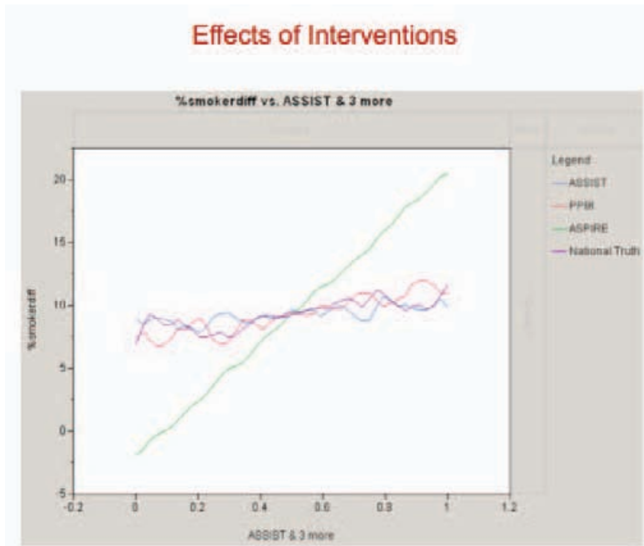



**Figure 5. Example Model Output: Relative Effects of Intervention Programs**

The other interventions, in this example PPBI, ASPIRE, and National Truth Campaign, are more effective then the ASSIST intervention program at a level up to 50% population coverage, however, the ASSIST intervention then dominates the other interventions. Using this example, if a policy maker only had enough funds to invest in a program that influenced up to 50% of the population, we would recommend that they chose any of the programs except the ASSIST program. However, if they had enough funding and a desire to implement an intervention program over more than 50% of the population, then the ASSIST program is a much more effective choice for reducing population smoking rates.

As part of our model assessment, we decided to compare the performance of our agent based simulation model in predicting smoking rate reduction vs. how the odds ratios predicted the same outcomes. Figure 6 shows a comparison of how the simulation output (plot points on the graph) compare to our estimate of population smoking reduction based on the odds ratios (linear line plot). Most of the odds ratios performed as expected with the exception of the Truth Campaign intervention, which led us to hypothesize, that the original research to support the effectiveness of this intervention may not have considered the effects of social

---

[2] See software downloads at http://harvest.nps.edu/

influences on the outcome. This area is one that we will explore with further research.

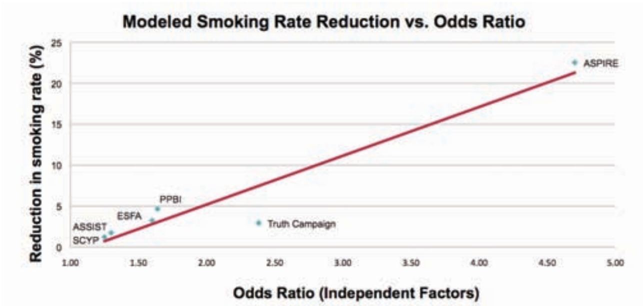



**Figure 6. Example Model Output: % Smokers Pre and Post Interventions**

Next, we explored how well the agent based simulation model showed how the population behaved over time based on the human and intervention program characteristics.

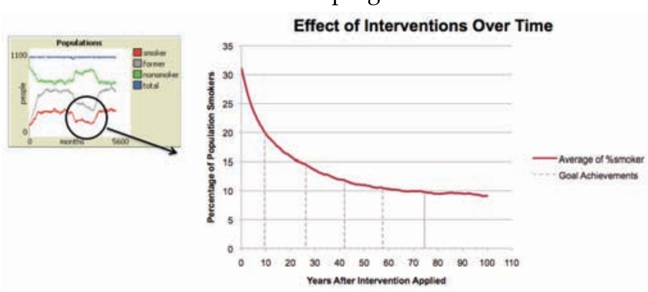



**Figure 7. Example Model Output: Time for Interventions to take effect**

Figure 7 shows how the model behaved for an example combination of interventions. In this case, the model projected approximately 60 years to reduce the population smokers by 20%

## CONCLUSION

Leading into IDFW 20, Team 8's objectives centered on developing a robust experimental design and conducting verification and limited validation of the agent based simulation model that we developed using NetLogo©. During the workshop, we accomplished all of these objectives and realized tremendous improvements through the help of solid preparation, other IDFW 20 participants, and the expertise of the SEED Center for Data Farming professionals.

Special thanks goes to Santiago Balestrini, another workshop participant, for selflessly providing his time and NetLogo© knowledge to help improve our model run time. Through improvements to our experimental design and agent based simulation model, we are now able to explore a more robust sample space with ½ of the original model run time resulting in a more robust analysis capability.

Our initial technical observations as a result of this workshop include:

• Validation that the influences of a social network are important to consider when evaluating the effectiveness of intervention programs on reducing population smoking rates

• An estimate for the length of time each intervention program or combination of programs need to be funded to ensure effective reduction in population smoking rates

• An understanding of which type of intervention(s) to invest in based on the size of the population that can be reached based on limited funding.

Following this workshop, CTC will build on the insights gained during IDFW 20 by continued enhancement of the agent based simulation model. This work supports a larger research effort to support policy decisions that effect funding for different types of intervention programs based on expected effectiveness in reducing smoking rates with extension to disease prevention.