Project Report
ACTA-6

Development and Validation of the Public-Facing SimAEN Web Application

J.D. Alekseyev
M.R. Chmielinski

1 August 2022

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

This material is based upon work supported by the Dept of Health and Human Services under Air Force Contract No. FA8702-15-D-0001.

This report is the result of studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This material is based upon work supported by the Dept of Health and Human Services under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Dept of Health and Human Services.

Massachusetts Institute of Technology
Lincoln Laboratory

Development and Validation of the Public-Facing SimAEN Web Application

J.D. Alekseyev
Group 23

M.R. Chmielinski
Group 21

Project Report ACTA-6

1 August 2022

This material is based upon work supported by the Dept of Health and Human Services under Air Force Contract No. FA8702-15-D-0001.

Lexington Massachusetts

This page intentionally left blank.

ACKNOWLEDGEMENTS

The authors are grateful for the dedicated efforts and expertise of all who contributed to the development, execution, and validation of the web application: Lauren Finklea, Jody McLean, Curran Schiefelbein, Robert Hallowell, Dieter Schuldt, Greg Gianforcaro, Rick Gervin, Sarah Miller, Joaquin Avellan, Bhavani Ananthabhotla, Nathaniel Hanson, Maegan Jong, Bill Streilein, Adam Norige, Brent Cassella, and Elizabeth Bernardo. We are additionally grateful to our validation testers for taking the time to meet with us and share their experience. We thank Elena Britos and Chet Beals for their assistance in preparation of the report.

This page intentionally left blank.

TABLE OF CONTENTS

This page intentionally left blank.

LIST OF ILLUSTRATIONS

This page intentionally left blank.

LIST OF TABLES

This page intentionally left blank.

1. INTRODUCTION

During a pandemic such as COVID-19, non-pharmaceutical interventions (NPIs) can help protect public health; however, it is not always clear which actions will have the greatest positive impact, or what the trade-offs are between different options. Exposure Notification (EN) was introduced as a prevention measure during the COVID-19 pandemic to supplement traditional contact tracing activities [1]. To predict the estimated impacts of EN, a model for "simulation of automated exposure notification" (SimAEN) was developed by researchers at MIT Lincoln Laboratory (MIT LL) with CDC funding [2]. The model was published through an accessible web interface, available for use by the general public at https://SimAEN.philab.cdc.gov/.

Initial efforts to engage with different U.S. stakeholders resulted in very little feedback about the underlying model's design and utility. The model development team presented its approach, inputs, initial interface, and analysis results to a group of state representatives, as well as individually to a handful of analysts working for various states. Reviewers struggled to comment on or engage meaningfully with the model through the initial interface. Results from past technology transfer efforts suggested that when users are presented with technology they cannot connect to their existing mental models, they struggle to provide feedback. The user experience (UX) team members hypothesized that reviewers were unable to meaningfully connect with the application, even though a number of them were analysts familiar with similar tools and workflows. Because consistent feedback from CDC sponsors centered around transition to a public audience, the application would need to be usable by non-specialists as well as subject matter experts.

With encouragement from CDC program stakeholders, the UX team met with model developers and interface developers to translate the model build into a web interface aimed at showcasing the model and eliciting feedback. The model was then reviewed in usability sessions with a number of state and federal public health decision-makers and epidemiologists. This led to several model adjustments, along with substantial questions regarding which parameters the model should expose and which metrics the model should calculate and display. This paper discusses the work MIT LL performed to translate the model into the web interface, and the substantive feedback this effort gained from the intended user community.

This page intentionally left blank.

2. BACKGROUND: USER EXPERIENCE WITH TECHNOLOGY

Whenever users open a computer application, they are presented with a screen and a series of selections (the interface), which requires them to form goals, translate their goals to the interface provided, and interpret the feedback provided in the interface, shown in Figure 1. For SimAEN, the web interface maps some elements of the real world to a mathematical model [2]. That mathematical model is translated into computer code [3]. The interface provides a way for users to set parameters within the code, run the model, and display outputs.

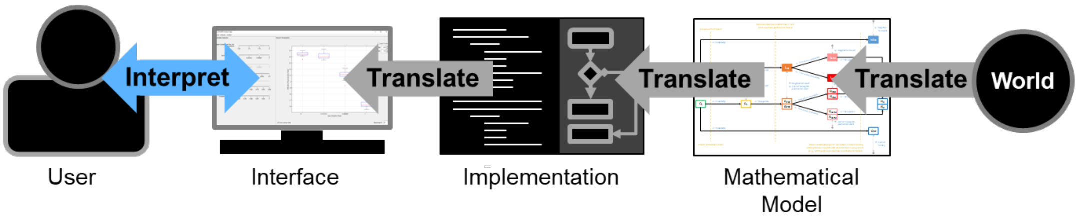

Figure 1. Translation and interpretation chain between the real world and technology users.

Users experience technology through the interface, and for many users, "*the interface is the product"* [4]; the underlying code and models are not visible. Users search the interface for clues on what to do or how to interpret the information provided. When interface selections match an unfamiliar mathematical or implementation model, users can experience difficulty using the software [5]. As discussed in [6], this difficulty is also found when the problem domain understanding and assumed workflow inherent in the interface are understood by the development team, but not in terms or workflows familiar to the intended user community. As software becomes more ubiquitous in people's lives, many companies turn to sleek user experience and even gamification to win over users. In turn, users' expectations are growing while they become less tolerant of software that does not meet their expectations [7]. Complicating matters further, people's emotional, visceral responses can override difficulties they experience during use of a product [8]. Many users will tolerate clunky use if they have a pleasant visceral response to the interface.

The technology developers, then, are obliged to take a step towards their users and detail reasonable examples of goals, show potential users how to map those goals to the new technology, and explain how to interpret the output. Undertaking this work helps potential users understand the use of the interface and technology in a functional way, which can improve user acceptance of the existing tool, and/or identify modifications that make the tool more useful, and/or identify other communities the tool could benefit. Presenting a product in a pleasing way also signals to users that the development team cares about their needs, and that they are capable of producing a positive experience with their technology. A pleasing and engaging interface can also build early trust with project sponsors and users alike, which can afford the team the time and flexibility that may be needed to adjust for a new user community.

This page intentionally left blank.

3. INITIAL SIMAEN INTERFACE

SimAEN uses a simplified version of real-world actions surrounding COVID-19 exposure and interactions with public health, and translates those into an agent-based model. With agent-based models, a number of runs can be performed to see how results vary due to diversity inherent in population-level interactions (often implemented probabilistically). This allows for emergent behaviors and a more realistic picture of the interactions between policy, workflow, and disease spread. Each agent in the model represents a person, with the model governing interactions and likelihoods on a personal level. As shown in Figure 3Figure 2, the underlying simulation advances one day at a time, and starts with a small number of infected people. Parameters govern how many people the initial infected person contacts, and the likelihood the contact will become infected, will get tested, will get contacted by public health, and so on. Each person in the model updates their status each model "day," with a percentage getting tested, processed by public health, spreading infection to others, and/or quarantining (see Program Report ACTA-5 [2] for a detailed explanation of model parameters and function).

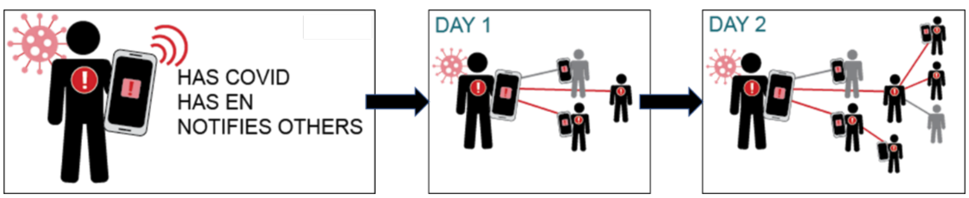

Figure 2. Notional step-through of SimAEN model function.

The initial interface was built in MATLAB software as shown in Figure 3. SimAEN has over 50 input parameters that can be adjusted to match conditions the user desires to model. The *About* tab was used to expose a small number of parameters to users (*Parameter Selection*, at left), each with a number of settings, and to present model results (*Results Visualization* panel, at right). This example shows *Probability of Detection* set to 1, which means that we are telling the model that each EN app will detect another EN app when it should. The *Results Visualization* panel shows that, given these input parameters, the *Effective Reproduction Rate* of the disease decreases as app adoption increases. Additional parameters can be found in the *Analysis* and *Control* tabs at the top, and additional settings and selections are found in the bottom left, bottom right, and top right around the *Results Visualization* panel.

When presenting results, the modeling team typically varied parameter inputs in a systematic fashion and created a series of plots shown together, as shown in Figure 4. In the image, each chart has one

Probability of Detection value shown above the plot, and four different *App Adoption Rate*[1] values below. The interpretation in the green callout box was determined by looking at the pattern of values from left (0.09/9% *Probability of Detection*) to right (1.0/100% *Probability of Detection*) on each chart, then at the pattern differences between charts; the more accurate EN is in detecting "close contacts," the more disease spread is suppressed at higher app adoption levels. While this may appear to be common sense, simulating these results can help jurisdictions explore tradeoffs and identify inflection points for the levels of adoption and the EN sensitivity and specificity in their regions.

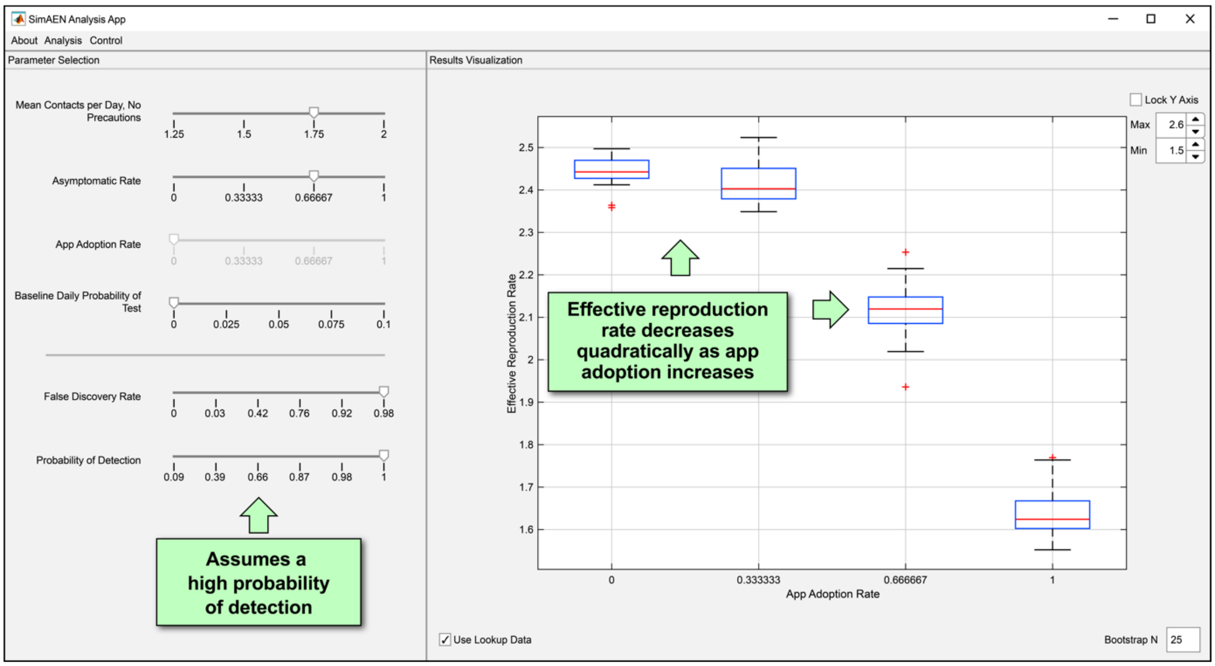

Figure 3. SimAEN MATLAB interface with descriptive callouts.

[1] "App adoption" in SimAEN refers to the decision by an individual user to enable the EN capability on their own phone; "app adoption rate" describes how widespread EN usage is in the region. In later iterations of SimAEN, we switched to the term "EN adoption rate." This paper uses "deployment" rather than "adoption" to refer to the decision by a public health authority to provide EN within its jurisdiction.

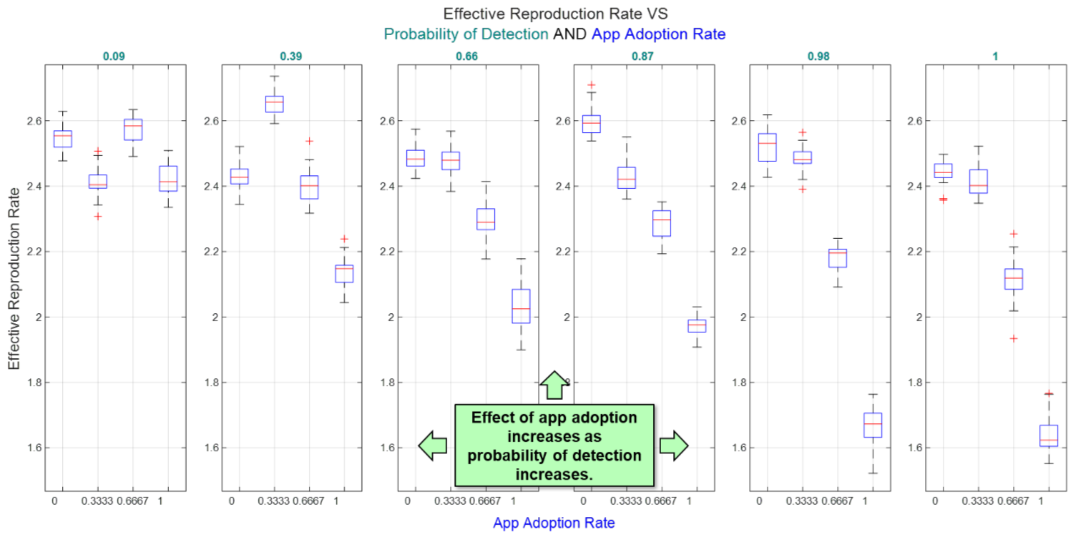

Figure 4. A number of SimAEN outputs and comparative analysis interpretation.

The modeling team presented the model motivation, scope, SimAEN modeling workflow, and an example analysis/experiment to a group of program stakeholders. The presentation did not result in any questions or follow-on conversations from the audience. The team additionally sought out and met with a number of state representatives who were open to new analytical approaches one-on-one. Comments from state representatives ranged from "looks good" to "I just need to play with it a bit to get a feel for it." This set of comments did not provide constructive criticism that would help the modeling team improve and iterate on either the model or interface implementations.

4. BUILDING AN APPROACHABLE WEB APPLICATION

The user experience (UX) team members hypothesized that initial reviewers were unable to meaningfully connect with the application, even though a number of them were analysts familiar with typical analysis processes and workflows [6]. Additionally, stakeholders at the CDC indicated they would like to host the model on a CDC-affiliated website, but that the interface would need to be understandable and usable by non-experts. In order to present the model to a general audience and connect with users, as well as to meet CDC requirements, the UX team needed to identify our users' requirements, understand the model functions, and map between them. Our primary goal was to provide a low-barrier entry point for people to begin to understand model functions, form their goals, and use the SimAEN model to meet them. Because CDC staff were acting as our user representatives, we needed to earn their buy-in in order to host the model. In turn, the expectation was that follow-on usability sessions with federal and state end users would help guide the team and the CDC to refine model requirements and further outreach.

4.1 NOTIONAL SIMAEN USER PERSPECTIVES

Initial work to present SimAEN to new users focused on developing a hypothesis of who our users were, and what their priorities were. We could then tailor the web app to the needs of these users. Initial user modeling could also help tease out the modeling team's implicit assumptions about users, which could be tested to improve team understanding. At the time, a number of states were considering EN, but had yet to decide whether or not to adopt the service; it was a new technology, and a number of states did not know what to expect from enabling the service. Aligning with the interaction design process described in [5], we created a persona of a likely main user:

> **The Advocate *(Jerry)*:** *With COVID, we're constantly feeling a step behind. It's been nothing but a scramble since the U.S. started reporting cases. My department has been keeping an eye on these smartphone apps for a while—a lot of people in my department felt like they were at best some toy for the technologists, or were afraid that researchers were trying to replace our contact tracers. Most of our managers have been in "wait and see" mode before wanting to spend the money to build the infrastructure. As someone that has grown up with technology and uses my smartphone for everything from ordering groceries to checking the latest news articles, the idea of using it to help stop the spread of COVID-19 is really enticing. I put a lot of reports together for my leadership on other matters, and see a lot of promise in this tech, but EN questions have been hard to answer and the budget is always stressed. I'd love to be able to answer the "what ifs" that come up with the people who handle the budget and can make EN a reality in my state—if we turn on the service, will people overwhelm our testing facilities? Will we catch more cases than we would with contact tracing alone? Just what can we expect from turning this service on?*

We also created a narrative describing "magic" use of the software:

> *While looking through the CDC tools for public health, Jerry comes across a tool that helps state public health workers like him evaluate NPIs, including EN. He easily accesses the web application. The interface looks appealing and approachable. He reads a bit about the model, and determines that it looks well-researched. He quickly learns that it can help him compare different strategies for NPIs, so he decides to give it a try. He quickly learns what the parameters are for, selects a few, runs the model, and compares results in a few minutes. He determines that this is what he needs to help his team decide about EN use, so he shares the link with a few colleagues.*

From this exercise, the UX team was able to approach the model with a specific notional user and specific investigative questions in mind, thereby to develop application requirements:

- If a state deploys EN,
    - How many new COVID-19 cases can they avoid?
    - What is the impact on testing facilities and wait times?
    - How many additional contact tracing calls will this generate, or require?
- Requirements:
    - As a novice user, I need
        - high-level information about model function to determine whether it will meet my needs
        - to get up to speed quickly with model parameters to understand how to match the model to my situation
        - a polished interface to reassure me that thought and consideration has gone into the model build
    - As an advocate for EN in my jurisdiction, I need
        - to compare use of EN to non-use, through model output metrics
        - easy-to-understand data visualizations that help me share model outputs with others

4.2 UNDERSTANDING SIMAEN

In order to help our potential users understand the model and build test scenarios, the UX team first needed to understand its inputs and outputs, as well as how model developers and analysts on the project team were using the model to produce and discuss results.

MATLAB interface walk-through. We started with a walkthrough of the interface with the model developers shown in Figure 3 to understand for ourselves how to set and run the model. In order to set up the experiment to determine that the "effective reproduction rate decreases… as app adoption increases," users not only needed to understand what the six parameters shown in the interface were doing, but also what the other 44 parameters were; this is knowledge the model development team had, but which new users would need to learn. In the example shown, they needed to know that the *Probability of Detection* parameter represented the concept that, given two people in proximity with each other with EN enabled, there is some probability that the Bluetooth messages will be received by the "close contact" smartphone. Additionally, they needed to understand that the slider value range of [0.9, 1.0] corresponded to the range of probabilities from 9% to 100% in common parlance. Finally, in order to use this in a meaningful way, the application user needed to understand what could be a "reasonable" setting for this value in the relevant region or state. These hurdles were equally relevant to interpreting and using other regional parameters (*False Discovery Rate*, *Mean Contacts Per Day*, etc.), as well as *Bootstrap N* in the bottom right of the panel (i.e., the number of model runs which would be aggregated to provide outputs) and the Y axis values in the top right of the panel. These settings had meaning for the model developer, but likely did not have meaning for the novice user, who would need to learn how to set the parameters and interpret the results.

MATLAB analysis walk-through. To understand the type of outputs that were reasonable for the model, and how to interpret the results, the application developers and the UX team also worked through output graphs from the application, such as the sample shown in Figure 5. Here, the charts show how the daily mean (blue) over a number of runs (all in gray) aligns with numbers expected from research or experience (red). *Counts* refers to the raw number of people in the simulation, and *Day* is the number of simulation days (100). Comparing the lower *Mean Interaction Rate* (2.1, top row) with the higher rate (2.5, bottom row), we can see that the lower interaction rate resulted in a lower rate of infection over the 100-day run. Comparing the "EN" plots (right-hand column) to the "Baseline" plots (left-hand column, i.e.), the deployment and user adoption of EN appears to depress the number of new infections over the course of the 100-day simulation. Additionally, in the top right corner, we see that EN use in conjunction with the lower *Mean Interaction Rate* (2.1) seems correlated with a relative stable amount of viral spread from days 50–100 compared to the baseline.

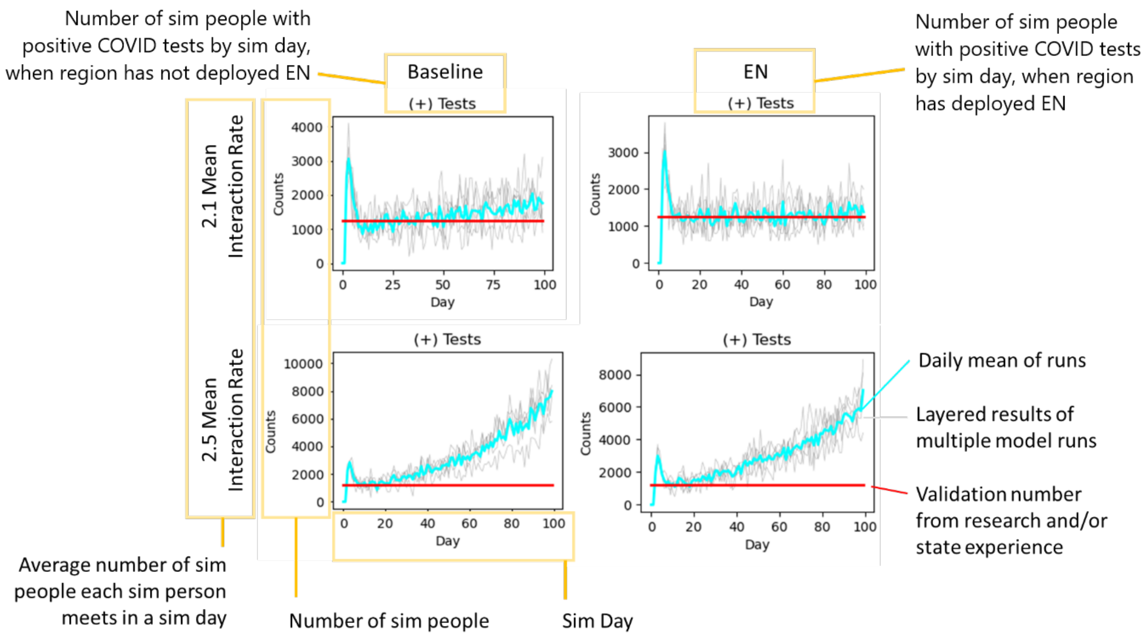

Figure 5. Model output graphs and explanations.

5. WEB APPLICATION USER EXPERIENCE DEVELOPMENT

From the work undertaken to understand our users and the model, it was clear that there were a number of parameters that could be set "under the hood." For instance, the modeling team knew what a reasonable number of runs would be to smooth out stochastic effects, and also had knowledge of what reasonable performance to expect from EN deployments (i.e., what *Probability of Detection* and *False Discovery Rate* values were predicted for real-world EN deployments, and why these are inversely related). They were also more likely to want to set and vary EN *Probability of Detection* and *False Discovery Rate* parameters. Our users, however, were more likely to want to turn EN on or off to model choices they were likely to make, or to prefer a single setting that incorporates the relationship between *Probability of Detection* and *False Discovery Rate* and excludes nonsensical choices.

As shown in Figure 6, the team worked through a number of iterations with the modeling team and CDC user representatives to determine a reasonable number of input parameters, the default value set for all parameters, and user-focused labeling, as well as to develop an initial view of metrics that would be more intuitive for our users to understand and explore. Finally, the UX team considered the information requirements that new users would have, namely to "get it—what it is and how to use it—without expending any effort thinking about it" [9]). The closer our team could get to the ideal of a large group of users seeing the site and being engaged enough to try the model and have conversations with our modeling or UX team, the more successful our program would be.

After selecting the variable parameters and determining the output display tasks, the team built out the full intended workflow and detail documentation as shown in Figure 7. As desired by the sponsor, the website implementation team used the automated Web Accessibility Evaluation Tool (WAVE) to identify areas that needed to be brought into compliance with Section 508 standards for accessibility and usability [10] [11].

Figure 6. The iterative design process, moving from wireframes and content mockups (left) to polished design and content (right).

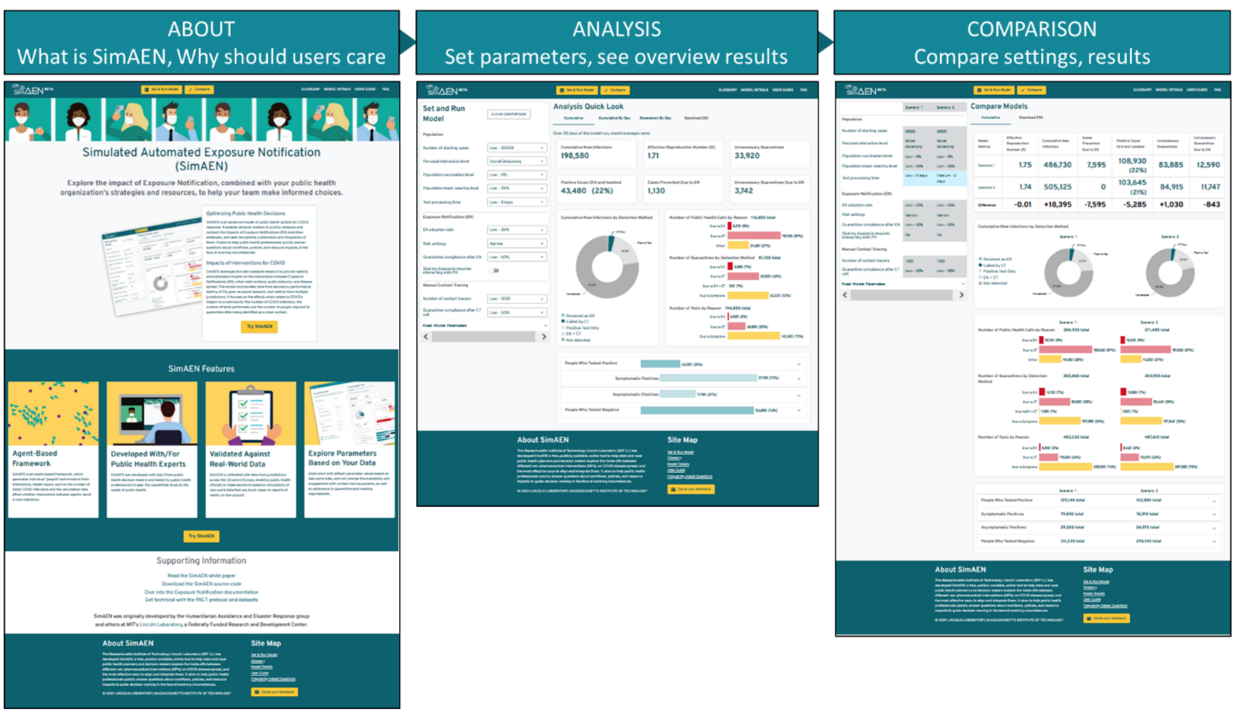

Figure 7. Basic web app intended workflow at top, with actual application screens and content.

First, users would land on an ABOUT page, which provided high-level details about the model, its focus, and its main features. At the bottom were links to additional resources for researchers, such as reports produced by the modeling team and the open-source model code. Most buttons on the page would direct users to the ANALYSIS page, where users could very parameters and see immediate results. Also on the ANALYSIS page, users could create two different runs of the model to compare, viewing results side-by-side on the COMPARISON page. Walking through this workflow would allow users to understand some basics about the model, play around with model function to get an idea of what outputs the model produces, and compare two different settings for a more in-depth analysis.

The team also built user support functions into the site as shown in Figure 8, to support users from novice through more experienced analysts. Users who need additional support could look to the USER GUIDE to walk through the site functions, or visit the GLOSSARY listing for each parameter to read what each does (short "tooltip" descriptions are also accessible from the ANALYSIS page by hovering over each parameter field). Those looking for more details of how the model was built, how it runs, and how the team validated its performance could read the MODEL DETAILS page.

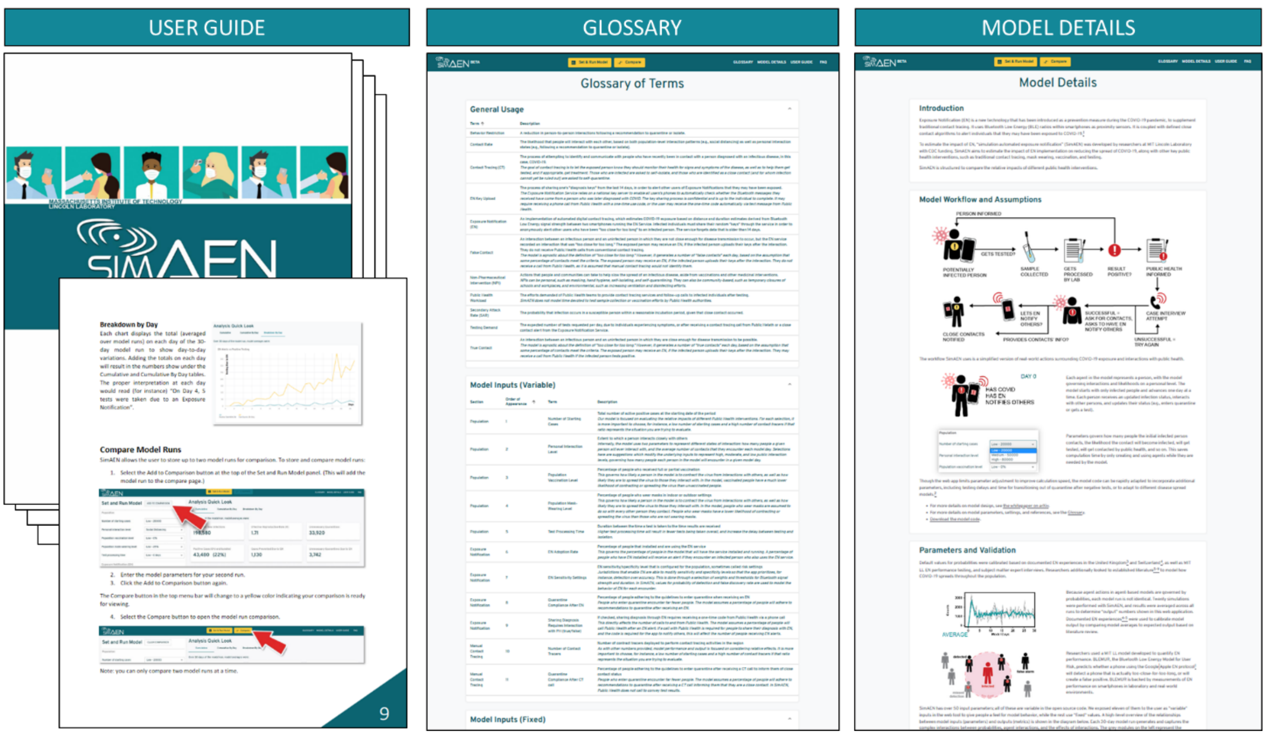

Figure 8. User support features screens available from the app menu bar.

This page intentionally left blank.

6. USER OUTREACH AND ASSESSMENT

6.1 OBJECTIVES

A core principle of good website design is to test assumptions about what is a reasonable workflow and information presentation for end users, *with actual end users*, in order to identify usability issues and mitigate them. Conducting these assessments helps the development team to gain a better understanding of user perspective and needs, and to iteratively improve the product. Once the web tool was live and accessible, our primary objective in conducting user outreach was to investigate users' reactions to the look and feel, information presentation, and workflow, in order to validate or refine the website implementation as well as our understanding of what information users need. We also wanted to identify whom users thought the site was geared towards, in order to validate or refine our assumptions about who our audience should be.

6.2 METHODOLOGY

We developed a script for demonstrating site functionality to each user, and a standard series of questions to focus user comments on aspects of interest. We wrote a simple analytical scenario to orient attendees to the expertise and concerns of our assumed user. The scenario could be walked through and described succinctly:

I'm a state public health decision maker, and I want to understand what to expect once we roll out EN. Let's check on what the outcome of that policy change may be.

Seven individual sessions were held over teleconference, as travel and in-person meetings were restricted due to pandemic conditions. Each session was scheduled for 45 minutes. The session moderator shared their screen to display the live SimAEN website, read from a script to ensure consistency, and walked through a portion of the workflow, stopping to ask a series of standard questions on each screen (see full moderator script and questions in APPENDIX A):

1. **Landing page (ABOUT page):** Short description of motivation of the model, and how the website was hosted. Questions concerned initial site impression, as well as information helpfulness.
2. **Set and Run Model – parameter settings (ANALYSIS page):** Walk-through of setting parameters to investigate the scenario. Prior to discussing information visualization, questions concerned usefulness and completeness of the parameters.
3. **Set and Run Model – data visualization (ANALYSIS page):** Interpretation of visualization display according to scenario. Questions concerned usefulness and completeness of the visualizations.

4. **Comparison visualization (ANALYSIS to COMPARISON page):** Walk-through of rationale and setup of comparison display. Questions concerned the usefulness and completeness of the visualization.

At the completion of the walk-through, the System Usability Scale (SUS) questions [12] were given to attendees.[2] Discussions, both verbal and in-software chat, were recorded by the session moderator. Qualitative analysis was conducted on the comments across attendees, and a breakdown of SUS scores were calculated where collected. Analysis of question answers focused on whether users indicated that this site was "for them," or whether they thought it was geared towards someone else; whether the information density was appropriate; and how understandable the information was, as presented.

6.3 RESULTS

As shown in Table 1, seven validation testing sessions were held to discuss the app with 13 people. Session attendance ranged from one-on-one sessions (marked "S" for solo), to a group session with five attendees (marked "G" for group). Sessions ranged in time from 45 to 60 minutes, depending on the length of discussion. All attendees worked in public health departments, and there was a split of feedback from people that worked at the state level (8), and those that were employed at the federal level (5). Four attendees indicated working as an epidemiologist, and four indicated working in some supervisory capacity. One attendee indicated having both aspects to their work. There were an additional four attendees that worked for IT departments, or in a computer science capacity. One attendee indicated working in quality control, while one did not provide a work function.

[2] The SUS was not administered to one group session consisting of five (5) attendees due to time constraints; it was administered in all other sessions.

Table 1: Overview of Validation Session Attendees

	Participant Number												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Session	G1	G1	G1	G1	G1	S1	S2	S3	G2	G2	G3	G3	S4
State Representative	X	X	X	X	X			X	X	X			
Federal Representative						X	X				X	X	X
Epidemiologist		X		X			X		X				
Supervisor/Manager	X		X				X	X					
IT/CS										X	X	X	X
Quality Control					X								
Not Indicated						X							

G = Group session; S = individual session

We issued the System Usability Scale to attendees P6-13, as shown in Table 2 (see APPENDIX B: System Usability Scale for question text and full grading scale). The SUS is a common tool used by UX practitioners to gauge the usability of their sites, which provides 10 common questions for users to rate their agreement from a scale of *1-Strongly Disagree* to *5-Strongly Agree*. Part of its appeal is the ability to "grade" a site with reference to scores from other public applications, and get a sense of how usable the site is. Given the grading scale published by Lewis & Sauro [13], a score of 84.1 or above translates to an A+, or better than 96% of public sites measured: 2/8 of our attendees gave the SimAEN web app an A+. However, a score of 51.5 or below translates to an F, or better than only 14% of public sites measured; 4/8 of our attendees gave the web app a score of 50 or below. The qualitative analysis conducted on attendee responses shed some light into the wide variation of responses and reported experience, though a wide range of scores is not unexpected for the first round of testing with intended end users.

Table 2: System Usability Scale Scores Across Attendees 6–13

	Epi	Mgr	IT+	NI	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
P6				X	5	2	4	2	5	1	4	1	4	1	**87.5**
P7	X	X			4	3	4	4	4	1	3	2	3	2	**65.0**
P8		X			4	1	5	1	5	1	5	1	4	1	**95.0**
P9	X				1	1	3	1	2		4	4	1	5	**50.0**
P10			X		1	1	3	1			4	4	1	5	**45.0**
P11			X		4	4	2	5	4	2	3	4	2	3	**42.5**
P12			X		3	4	2	5	3	4	3	4	2	4	**30.0**
P13			X		4	2	5	1	3	2	4	2	3	2	**75.0**

Av. **61.3**

Epi = Epidemiologist; Mgr = Manager or Supervisor; IT+ = IT or Computer Science; NI – Not Indicated

As shown in Table 3, though 7/13 of our attendees were able to identify where to take action on the ABOUT page, and 4/13 specifically approved of the general look and feel of the site, 3/13 had basic questions about the intent of the site, and 4/13 had basic questions about what the SimAEN model is, or does. Additionally, 3/13 had trouble finding or understanding the buttons presented, and 5/13 had questions about terms. One attendee indicated there being too much information on the page, while 2/13 wanted to see more detail about what the model is, or does. Comparing responses from P9 and P10 in Table 1and Table 3, both attendees indicating wanting more detail were in a session with an epidemiologist.

While discussing the ANALYSIS page, 3/13 attendees specifically mentioned liking the look and feel, and 4/13 indicated that the page contained too much information or was too busy, with the one attendee from the ABOUT page being joined by three new attendees remarking on information density. However, one participant indicated that the information was much more understandable once the moderator walked through use, and three additional attendees remarked that the selections made sense, including two of whom felt there was too much information presented. Additionally, 5/13 indicated that they would want to play around with the selections themselves to get a feel for what is happening. This likely indicates that the content of the information was reasonable, but there were either too many selections provided on the screen, or they were presented in a confusing manner. Considering that 3/13 wanted to use “hard” numbers, 4/13 wanted the meaning of each parameter, and 5/13 had specific questions about how the model handled different situations, there was likely a more refined drill-down of information required in order to best balance between casual use, and the more detailed requirements of those wanting to dig deeper.

Once on the COMPARISON page, 3/13 specifically mentioned liking the comparison shown, while 2/13 tried out their understanding of the metrics provided. In general discussion, 3/13 indicated both that they would use it if it were their job to, and that it was not their job to use the site. Only one person specifically mentioned wanting to use it personally.

Throughout the sessions a number of attendees indicated who they thought would want to use the application, though these indications were inconsistent from user to user. Where one attendee mentioned that it would be good for executive use, another mentioned that decision-makers in their jurisdiction would need much "harder" numbers, or the attendee would get "blown off" trying to present model results to them. Additionally, while one attendee noted that they thought epidemiologists would love the interface, another thought that epidemiologists would need a more detailed interface—neither of these comments came from an epidemiologist, however. Two epidemiologists did indicate wanting "hard" numbers, and had questions about how aspects were presented in the model, indicating that there at least needed to be more detailed information available to them to browse and understand. One epidemiologist commented that "thinking back to last August, it would have been hard [to make decisions] because of lack of data; something like this could have been helpful at a very high level to determine 'do we want to do this or not,'" providing an indication that the model would be a beneficial one in the right context of use. Another attendee commented that the app would be good as a public relations tool, while another indicated that it would be good as a "story telling" tool from the public health perspective. Yet another attendee mentioned that the app could be useful for smaller jurisdictions without their own embedded modeling team.

Table 3: Specific Mentions by Attendees and Site Page

	Participant Number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
ABOUT page													
Knows where to take action							X	X	X	X	X	X	X
Approved of general look and feel	X						X	X					X
Questioned site intent		X			X	X							
Questioned what SimAEN is	X				X	X							X
Trouble finding/understanding buttons		X			X							X	
Trouble understanding terms	X					X	X				X		X
Too busy/too much information						X							
Wants more detail									X	X			
ANALYSIS page													
Likes the look					X	X		X					
Too busy/too much information						X	X				X		X
Easy to understand once broken down									X				
Selections make sense						X	X	X					
Would play around to learn							X	X			X	X	X
Good for executive use										X			
Hard to use to report to decision makers									X				
Good for general public use										X			
Good for Epidemiologist use						X		X					
Note detailed enough for Epidemiologist										X			
Wants "hard" numbers					X		X		X				
Questioned numbers used					X	X			X				
Wants meaning of each parameter					X	X	X	X					
Had modeling specifics questions					X	X	X		X	X			
COMPARISON page													
Likes comparison	X					X		X					
Felt confident interpreting						X							X
Would have someone else interpret									X	X			
Helpful for small jurisdictions									X				
Useful when there is no hard data										X	X		
Use to "tell a story" from Public Health								X					
Would want to share with leadership								X					
Good as a public relations app						X							
"Just play" as a private citizen							X						
General Comments													
Would use it "if it were their job to"								X			X	X	
Could see their org using (but not them)								X					
Could see themselves using							X			X			

7. DISCUSSION

During the sessions, attendees were not introduced to the User Guide, Modeling Details, or Glossary screens. It is possible that some of the comments would have been mitigated by this, or by allowing attendees to “drive” the analysis and visualization. However, there are a number of positive aspects: users generally understood the selections and the model once they were able to get the explanations they were looking for, and a number seemed interested in playing around with the interface further. The fact that some users wanted more detailed information while others wanted less information presented at once, which did not break out as expected along epidemiologist/non-epidemiologist user lines, indicated the user needs were also not so cleanly separated.

These observations suggest the direction of future enhancements to SimAEN's web interface, and considerations for the designers of similar systems:

1. **In order to broaden its usability and value, application updates would need to be considered from two distinct user information needs.** Some of our testers indicated finding immediate value in the application as-is; however, it was too detailed for some and not detailed enough for others. We will call them “The Strategist” and “The Analyst” here for ease of discussion.

    The Strategist. Some users wanted more context to understand groupings of selections, or comparisons they may want to try. One suggestion was to package the ANALYSIS page within the context of a strategy of intervention. One can imagine an overarching selection such as “Focus on Masking Initiative,” which could provide an analysis of how 50% vs. 75% compliance would affect spread. Additional context could be provided to model location or situational context, such a college campus, or a rural town.

    The Analyst. While some users wanted more high-level context, other users focused on the need for “hard” numbers and more details of model function. Additional drill-downs or detail information could be provided for these users, to allow them to trace parameters directly to how the model handled different interactions, what exactly was being modeled, and what assumptions were built in.

Addressing both users in one interface would allow even novice users to start at a high level and drill into the level of detail they need, but it would require a focus on understanding the details of information needs from each audience, and building the interface to support them.

2. **Ideal application development strategies would implement multiple testing cycles with end user communities.** Though we were unable to conduct a second round of testing and implementation improvements due to development and timing constraints (namely, developing an app quickly for a community in the midst of handling a pandemic), interactively adjusting user

understanding and application builds serves only to improve the end product. Asking a stressed workforce to learn new technology is difficult at best; we recommend carrying out testing and improvements during periods of reduced disease burden to secure community buy-in, and build broadly beneficial and accepted tools.

8. CONCLUSION

The work taken to translate the model to an accessible web interface is considered to be largely successful. Though not perfect, it allows users a way to explore model function in an accessible manner. The work undertaken to make the model more accessible resulted in actionable information for both broader outreach and usability improvements. In addition, the sponsor transferred the website and model to CDC servers for public use. We expect that this work will help to inform future modeling and development efforts for the public health community, and will serve as a use case of embedding User Experience considerations into any application build. The lessons learned from this work are broadly applicable to any situation where technology is being introduced to a new community.

This page intentionally left blank.

APPENDIX A: USER INTERVIEW SCRIPT

Hi [name]/All,

This session is intended to be an introduction to the SimAEN web tool, which provides a way for people to explore the model we built to investigate the relative impacts of non-pharmaceutical interventions available to public health. I plan to give a bit of a demo of the tool and get your feedback. Any questions or comments before we start?

1. Navigate to SimAEN website
    - **Landing page:** *SimAEN is hosted on a web page that will be accessible to the general public. On the main page, we have some high level information about the model and its intended uses.*
    - *The main take-aways here are that the simulation is intended to assist decision-makers in public health by allowing them to compare the effects of different strategies leveraging non-pharmaceutical interventions including exposure notification.*
    - *The benefits of using a model is that we can see the whole "world" we are measuring and can do what-if analysis for different scenarios and their impact—both of which are things we can't do in the real world. Let's see how this could work.*

    Prompt Questions:

    - What is your working background/job title (e.g., IT, epidemiologist, epidemiology)?
    - Before I move on, can I ask you/each of you to give me your impressions of the site at this point?
    - Do you have a sense of what SimAEN is, and what you would want to do with it?
    - Do you find the information on this page useful? Why or why not?
    - Is there anything missing from this page that you expected to see, or would want to see?

2. Click on "Set and Run Model"
    - ***Parameters:***
        - *We have a bunch of model parameters with different selections. These selection options are based the team's literature review of existing research surrounding COVID-19 spread, EN function, and participation in health interventions by the general public.*
        - *These selections are grouped into different headers based off of whether they represent real-world behaviors of the general population, how Exposure Notifications work and the options available for public health to adjust, and manual contact tracing (for comparison purposes).*
        - *Say I'm a state public health decision maker, and I want to understand what to expect once we roll out EN. Let's check on what the outcome of that policy change may be.*

- *I’m going to ignore the right side for now so we can focus on setting up our scenario.*
- ***First, I want to set my baseline to existing realities.** What I’m looking to do is compare the current situation in my state to one where EN has been rolled out, so I first want to represent the current situation in my state.*
- *And we have built in help like definitions for each of our terms.*
- *15000 active cases **(Leave at 20000)***
  - *Not too concerned about numbers not matching exactly, since SimAEN compares relative effects of different public health strategies, given what we know about the spread of COVID 19.*
- *Still have distancing recommendations (**Change to Medium)***
- *About 30% vaccination **(Set to Medium)***
- *Mask level **(Medium)***
- *Test processing **(Medium)***
- *EN adoption rate **(None – 0)***
- *MCT, manual contact tracing **(Leave at Low)***

3. **Analysis quick look:** The quick look screen is displaying the outputs for the 30 day run of the model at these settings.

Prompt Questions:

- Before I move on, can I ask you all what your impression is of this screen? Do you find the inputs and outputs understandable?
- Are there any inputs that you would want to see?

  - *There are a few things I want to point out.*
    - *Each number represents the **cumulative totals over the model run length**. That is something we set in the back end to maximize model performance.*
    - *The model estimates that after 30 days, we’re looking at something like 34,100 **cumulative infections**. While this is pretty high, the reproduction rate gives us a sense of how this might map to the real world. Remember that the model is estimating how many infections that may be unseen in the real world—the rates we are using are based off of published literature. Our model can “see” more in its world that we can in the real one. This high number is what experts estimate is actually happening.*
    - *The **“Not detected” number in the pie chart** shows this as well—we’re seeing here that EN, CT, and testing means that there is some sort of visibility into about 1/4 of the cases that exist in the model.*
      - *Asymptomatic spread*
    - ***Click through to each chart, discuss***
    - ***Add to comparison***

Questions:

- Is there any data presented here that doesn't make sense to you?
- Are we missing any inputs you would want to see? Any that we should remove?
- Any outputs that you would want to see? Any that we should remove?

4. **Set comparison state.** I want to now build my comparison. Though I can change any parameter I want, and all parameters if I choose to, I'm only going to vary ONE parameter here so that I can easily see the relative effects of my selections.

    - *Let's be really optimistic about EN adoption just for comparison.*

        - *I'll set my EN adoption rate to **high***
        - *I'll leave the other settings alone for now. If I want to compare settings, I can do that later.*
    - ***Add to comparison***

5. Click "Compare"

    - *On my comparison screen, I can see both of my selections side by side, with the parameter settings here with differences highlighted.*

Prompt Questions:

- What is your impression of the comparison screen?
- Is there any data presented here that doesn't make sense to you?
- Are we missing any inputs you would want to see? Any that we should remove?
- Are there any outputs that you would want to see? Any that we should remove?

This page intentionally left blank.

APPENDIX B: SYSTEM USABILITY SCALE

We'll now walk through the final questions. Please rate each on a scale of *1-Strongly Disagree* to *5-Strongly Agree:*

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

(as defined in [12])

Grade	SUS	Max-Min SUS	Percentile Range
A+	84.1 - 100	15.9	96 - 100
A	80.8 - 84.0	3.2	90 - 95
A-	78.9 - 80.7	1.8	85 - 89
B+	77.2 - 78.8	1.6	80 - 84
B	74.1 - 77.1	3	70 - 79
B-	72.6 - 74.0	1.4	65 - 69
C+	71.1 - 72.5	1.4	60 - 64
C	65.0 - 71.0	6	41 - 59
C-	62.7 - 64.9	2.2	35 - 40
D	51.7 - 62.6	10.9	15 - 34
F	0 - 51.6	51.6	0 - 14

Figure 9. Curved grading scale [13].

This page intentionally left blank.

GLOSSARY

CDC	Centers for Disease Control and Prevention (United States)
COVID-19	Coronavirus disease caused by the SARS-CoV-2 virus
CT	Contact Tracing
EN	Exposure Notification
IT	Information Technology
MIT LL	Massachusetts Institute of Technology Lincoln Laboratory
NPI	Non-pharmaceutical intervention (e.g., mask wearing, hand hygiene, quarantine)
SimAEN	Simulation of Automated Exposure Notification
SUS	System Usability Scale, a common UX tool to gauge website usability
UX	User Experience, focused on human factors related to technology use

This page intentionally left blank.

REFERENCES

[1] H. Bair, J. D. Wanger and N. R. Shah, "A Brief History of Exposure Notification During the COVID-19 Pandemic in the United States, 2020-2021," *Public Health Reports,* vol. 137, no. 2_suppl, June 2022.

[2] D. Schuldt, E. Londner, J. Saunders, M. C. Schiefelbein, R. Yahalom, A. Norige and W. Streilein, "The Simulation of Automated Exposure Notification (SimAEN) Model," MIT Lincoln Laboratory, ACTA-5, Apr. 2022.

[3] MIT Lincoln Laboratory, "GitHub - informaticslab/SimAEN: A predictive health analytics engine developed by MIT Lincoln Lab," 25 Aug. 2021. [Online]. Available: https://github.com/informaticslab/SimAEN. [Accessed 27 June 2022].

[4] J. Raskin, The Humane Interface: New Directions for Designing Interactive Systems, Addison-Wesley Professional, 2000.

[5] A. Cooper, R. Reimann, D. Cronin and C. Noessel, About Face: The Essentials of Interaction Design, 4th Edition, John Wiley & Sons, 2014.

[6] J. Alekseyev, "A Framework to Improve Evaluation of Novel Decision Support Tools," in *Advances in Usability, User Experience, Wearable and Assistive Technology. AHFE 2020. Advances in Intelligent Systems and Computing, vol 1217*, Springer, Cham, 2020.

[7] W. Gribbons, "The Four Waves of User-Centered Design," *UX Magazine,* 29 Jan 2013.

[8] D. A. Norman, "Emotional Design: People and Things," 3 Dec 2018. [Online]. Available: https://jnd.org/emotional_design_people_and_things/. [Accessed 24 06 2022].

[9] S. Krug, Don't Make Me Think! A Common Sense Approach to Web Usability. 2nd ed., New Riders, 2005.

[10] WebAIM at Utah State University, "WAVE Web Accessibility Evaluation Tool," [Online]. Available: https://wave.webaim.org/. [Accessed 27 June 2022].

[11] CIO Council Accessibility Community of Practice (ACOP), "Guide to Accessible Web Design & Development," March 2019. [Online]. Available:

https://www.section508.gov/content/guide-accessible-web-design-development/. [Accessed 27 June 2022].

[12] J. Brooke, "SUS-A quick and dirty usability scale," in *Usability Evaluation in Industry*, Taylor & Francis, 1996, pp. 189–194.

[13] J. R. Lewis and J. Sauro, "Can I Leave This One Out? The Effect of Dropping an Item From the SUS," *Journal of Usability Studies,* vol. 13, no. 1, pp. 38–46, 2017.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(01-08-2022)*	**2. REPORT TYPE** Project Report	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE Development and Validation of the Public-Facing SimAEN Web Application	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Jesslyn Alekseyev, Madeline Chmielinski	**5d. PROJECT NUMBER** 10383-2 **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108	**8. PERFORMING ORGANIZATION REPORT NUMBER** ACTA-6
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Centers for Disease Control and Prevention 1600 Clifton Road, Atlanta, GA 30329 USA	**10. SPONSOR/MONITOR'S ACRONYM(S)** CDC **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.
This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

During a pandemic such as COVID-19, non-pharmaceutical interventions (NPIs) can help protect public health; however, it is not always clear which actions will have the greatest positive impact, or what the trade-offs are between different options. Exposure Notification (EN) was introduced as a prevention measure during the COVID-19 pandemic to supplement traditional contact tracing activities [1]. To predict the estimated impacts of EN, a model for "simulation of automated exposure notification" (SimAEN) was developed by researchers at MIT Lincoln Laboratory (MIT LL) with CDC funding [2]. The model was published through an accessible web interface, available for use by the general public at https://SimAEN.philab.cdc.gov/.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		36	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18