Technical Report
TR-1287

Automated Contact Tracing Assessment

M.C. Schiefelbein
C. Ishikawa
E.H. Shen
R. Yahalom
A.S. Norige

17 November 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

Automated Contact Tracing Assessment

M.C. Schiefelbein
Group 46

E.H. Shen
Group 53

A.S. Norige
Group 21

C. Ishikawa
Kahuina Consulting

R. Yahalom
MIT Sloan

Technical Report TR-1287

17 November 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.

ACKNOWLEDGMENTS

Our sponsors at the Centers for Disease Control and Prevention, as well as the public health officers and health care providers throughout the U.S., and especially in Arizona, California, Colorado, Connecticut, Delaware, the District of Columbia, Hawaii, Massachusetts, Maryland, Michigan, Minnesota, Nevada, New Jersey, New York, North Carolina, Pennsylvania, Utah, Virginia, Washington, and Wisconsin, for sharing your expertise and insights with us during a time of tremendous crisis. We are deeply honored that you chose to collaborate with us on a novel approach to practicing public health.

The original PACT team, led by Marc Zissman, Ron Rivest, and Danny Weitzner, which helped establish this effort and provided insightful guidance through the entirety of the effort.

Our colleagues in the United Kingdom’s National Health Service, Ireland’s Health Service Executive, University of Maryland, GovTech Singapore, École polytechnique fédérale de Lausanne, University of Zurich, Fraunhofer-Gesellschaft, Robert Koch Institut, SAP, and the MITRE Corporation; your perspectives, experiences, and critical thinking enriched our work and encouraged our spirits.

The staff and leadership at Apple and Google, as well as at the Association of Public Health Laboratories, the Internet Security Research Group, the National Institutes of Health, and the MITRE Corporation, for undertaking the implementation and maintenance of a new kind of public health infrastructure across geographic, political, corporate, and digital borders, and for always being generous with your time and thorough with your considerations.

Numerous Lincoln Laboratory staff across all Divisions and Departments, for sharing expertise and resources, creative problem solving, and unflagging execution amidst unprecedented conditions. We especially thank Marc Zissman and Bill Streilein, for their guidance throughout the program, and Laura Glazer and Chet Beals, for their assistance in preparation of the report.

This page intentionally left blank.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Page

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

This page intentionally left blank.

1. EXECUTIVE SUMMARY

The COVID-19 pandemic placed unprecedented demands on the global public health systems for disease surveillance and contact tracing. Engineers and scientists recognized that it might be possible to augment the efforts of public health teams, if a system for automated digital contact tracing could be quickly devised and deployed to the population of smartphones. The Private Automated Contact Tracing (PACT) protocol [2] was one of several digital contact tracing proposals offered worldwide. PACT's mission—to preserve individuals' privacy and anonymity while enabling them to quickly alert even nearby strangers of a likely risky exposure—was adopted by Google and Apple and realized in the Exposure Notifications (EN) service and API [3] for mobile application development [4] [5].

The Exposure Notifications system, like many digital proximity tools, is based on Bluetooth signal strength estimation [5], and keeps much of the necessary information and computation on the smartphones themselves. It implemented a *decentralized* approach to contact tracing: the public health authority, and other governmental authorities, cannot access the records of an individual's encounters with others; nor is physical location used or shared by the service. Although the service is available on most modern iOS and Android devices, it is not enabled by default; the individual must opt in to use a particular region's implementation of the service, either by installing the regional app or by enrolling through a menu of regions in the operating system settings. Likewise, individuals must affirm their consent before the service can share anonymized infection status with the regional public health authority, and alert recent close contacts. The widespread availability of Exposure Notifications through Apple and Google's platforms has made it a *de facto* world standard. Determining its accuracy and effectiveness as a public health tool has been a subject of intense interest.

In July 2020, CDC's Innovative Technologies Team designated MIT LL and the PACT team as trusted technical advisors on the deployment of private automated contact tracing systems as part of its overall public health response to COVID-19. The Innovative Technologies Team sought to answer the following key question regarding automated contact tracing:

> *Does automated contact tracing have sufficient public health value that it is worthwhile to integrate it at scale into existing and evolving manual contact tracing systems?*

Rapidly rising caseloads necessitated parallel-path assessment activities of most mature systems at the time. When access to the Google and Apple Exposure Notifications system became available, MIT LL focused the assessment efforts on the systems being built and deployed. There were two immediate and significant challenges to observing and quantifying the performance of the system as a whole: first, the privacy-preserving design decisions of PACT and the system implementers denied access to system-level performance metrics, and second, obtaining accurate "ground truth" data about risky encounters in the population, against which to measure the detector performance, would require an unacceptable level of effort and intrusion. Therefore, MIT LL designed a set of parallel research activities to decompose the problem into components that could be assessed quantifiably (Bluetooth sensor performance, algorithm performance, user preferences and behaviors), components that could be assessed qualitatively (potential cybersecurity risks, potential for malicious use), and components that could be modeled based on current and emergent knowledge (population-level effects).

The MIT LL research team conducted early assessments of the privacy and security aspects of new EN app implementations and closely reviewed the available system code exercised by the apps, before conducting a series of phone-to-phone data collections both in the laboratory and in simulated real-world conditions. The data from these experiments fed into models and visualization tools created to predict and understand the risk score output of candidate "weights and thresholds" configurations for EN, i.e., to predict the performance of the system as-built against ground truth data for distance and duration of "exposure". The data and performance predictions from this effort helped to inform the global and local community of practice in making configuration decisions, and can help to predict the performance of future versions of similar tools, or alternative implementations of the current system. We conducted a human factors and usability review of early app user interfaces and messaging from public health, and designed a follow-on large-scale survey to investigate questions about user trust and system adoption decisions. The results of the human factors, user trust, and adoption studies were used by U.S. public health jurisdictions to make adjustments to public-facing communications, and were shared with Apple and Google to improve the user interface. Information gathered from public health experts enabled us to better understand conventional contact tracing workflows and data streams, and we incorporated that information into an agent-based model of "hybrid" contact tracing plus Exposure Notifications. We then combined it with emerging reports on vaccination, mask effectiveness, social interaction, variant transmissibility, and our own data on the sensitivity and specificity of the Bluetooth "dose" estimator, to predict system-level effects under various conditions. Finally, we helped to establish a network of Exposure Notifications "practitioners" in public health, who surfaced desirable system-level key performance indicators (implemented during 2021 and 2022, in the Exposure Notifications Private Analytics system, or ENPA).

At the conclusion of the program, many of the initial conditions of the pandemic had changed. The Exposure Notifications service was available to most of the world, but had only been deployed by 28 U.S. states and territories, and had not been adopted by much of the population in those regions. High case rates during the Omicron surge (December 2021 – January 2022) and newly available ENPA data offered the first hints at calculating "real" state-level performance metrics, but those data belong to the states and many are cautious about publishing. Although Google and Apple have stated that Exposure Notifications was designed for COVID-19, and will not be maintained in its current form after the pandemic ends, the public health and engineering communities show clear interest in using the "lessons learned" from Exposure Notifications and other similar solutions to preserve the capabilities developed and prepare better systems for future public health emergencies. The intent of this report is to document the work that has been completed, as well as to inform where the work could be updated or adapted to meet future needs.

2. PROGRAM OBJECTIVES AND TIMELINE

The overall effectiveness of an automated contact tracing system is affected by multiple factors: the sensor implementation, the risk estimation algorithm, the usefulness to individual citizens, the disease surveillance insights available to public health teams, and the population-level impact of the system on the spread of disease and other effects. Likewise, the purpose of the Automated Contact Tracing Assessment program (10383) was to take multiple parallel approaches toward accomplishing the following:

- Advise on system architecture, including both the technical design of private automated contact tracing systems, and deployment strategies for complete systems, as needed
- Assess the viability of Bluetooth Low Energy (BLE) or other approaches (e.g., ultrasound) for proximity detections, and work with public health partners to articulate what accuracy levels are needed
- Develop guidance for U.S. states on the most appropriate and effective ways of configuring and using the Google-Apple Exposure Notifications (GAEN, or EN[1]) service functions
- Assess the security and privacy of the implementations of the evolving EN service, both with respect to the implementation of the cryptographic protocol and the broader information privacy and system security considerations
- Provide analysis and advice about the integration of private automated contact tracing systems with existing and evolving manual contact tracing case management systems, and provide technical support to public health partners in assessing the best strategic choices for using systems such as EN
- Work with CDC and other recommended partners to evaluate the user experience design of exposure notification systems and related infrastructure for the purpose of identifying privacy, security, and usability best practices that will enhance public trust in the contact tracing activities
- Propose research studies that enable real-world evaluation of user experience as services are deployed, and undertake these investigations using human-computer interaction (HCI) methodologies, privacy policy expertise, and collaboration with other experts in the field

[1] "GAEN" and "EN" can be used interchangeably to refer to the Google-Apple Exposure Notification implementation.

- Work with CDC and designated U.S. state partners to design pilot projects intended to assess the efficacy of a private and effective exposure notification system, and provide technical advice during the execution of one or more state pilots.

Throughout the course of the program, the PACT members at MIT organized regular virtual meetings to bring together the global automated contact tracing community to learn from implementation experience, assess best practices, and provide focused feedback to Apple and Google, other system developers, and CDC. This helped to build a community of practice around automated contact tracing systems, and along with the MIT LL/CDC-hosted Learning Lab meetings for U.S. state partners, provided timely forums for the exchange of technical information and situational awareness of how various jurisdictions were making use of exposure notification technologies.

Figure 1 shows a high-level overview of the program efforts, in the context of major feature releases of the EN service, and of key pandemic events [7] [8] [9]. In the first five months of the program, EN was being "built while in flight": the first few jurisdictions (Austria, Germany, Ireland, the U.K., Pennsylvania, Virginia) had launched or were about to launch initial versions of EN-based apps, yet there was little to no data available to inform how to optimize the Bluetooth detector's configuration, and Google and Apple had just redesigned the risk scoring algorithm. System-wide interoperability concerns had not yet been addressed and questions of how to measure the public health impact had been deferred in the interest of building a minimum working system as fast as possible under emergency conditions. A team of staff from MIT LL, MIT CSAIL, MIT IPRI, and MIT Sloan worked with CDC, as well as with public health professionals across the U.S. and subject matter experts at the Association of Public Health Laboratories (APHL), Apple, Google, MITRE, Linux Foundation Public Health (LFPH), and the National Health Service (NHS), to reduce the uncertainty and close gaps in our collective understanding of how well EN works, and where possible, to make it work better for public health objectives in real time. After winding down the research-intensive period of performance, MIT LL continued to support the U.S. community of practice through hosting Learning Lab sessions and the publication of selected deliverables, and also continued direct support to one U.S. state[2], through June 2020. Throughout the course of the program, MIT LL presented work-in-progress and research results to the U.S. and international EN community, which contributed significantly to the two Risk Scoring Symposia and ImPACT 2021 "convening" workshop, and which continues to inform decision making around EN implementation and use today.

[2] The state has requested not to be identified, per the terms of the Non-Disclosure Agreement regarding pilot activities.

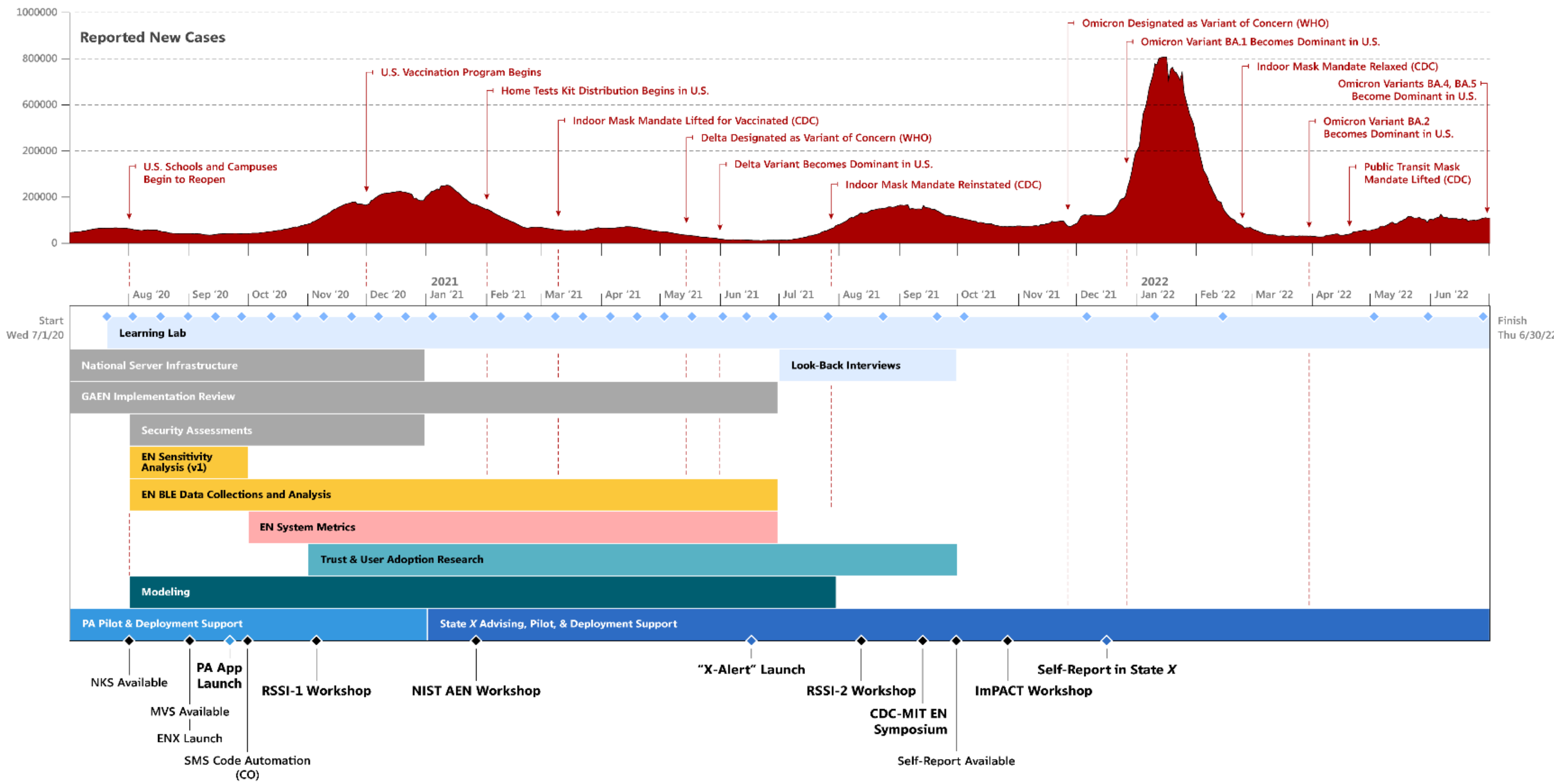


Figure 1. Timeline of Program 10383 period of performance.

3. BACKGROUND

It is well known that the SARS-CoV-2 virus can spread from person to person in the early stages of infection, before the infected person develops symptoms and becomes aware they are sick. Combined with the fact that the virus is easily spread through exhalations, it was imperative to quickly identify the "close contacts" of infected individuals and notify them that they may be carriers of the virus. Early in the pandemic, when the virulence of the disease was understood and before vaccinations, mask mandates, and other infection control mechanisms were available or widely deployed, public health officials realized an "army" of case investigation and contact tracing teams would be required to meet rising case loads. [1]

"Contact tracing" refers to the process of attempting to identify the "close contacts" of someone recently diagnosed with an infectious disease such as COVID-19, in order to quarantine and/or treat them. Case investigation, in which public health staff work with a patient to help them recall and identify close contacts, is one pathway by which contact tracing may be initiated. Another is when patients, on their own cognizance, identify and communicate the risk of infection to those whom they may have infected. "Automated contact tracing," then, is a novel third pathway, in which digital technology records potential exposures and communicates the risk of infection to close contacts, in order to augment the contact tracing efforts of public health teams. The term "automated exposure notification" refers to the same concept. MIT LL, along with collaborators in industry, academia, and health care, devised the Private Automated Contact Tracing (PACT) protocol to enable this novel approach.

Apple and Google implemented the PACT protocol in April 2020, and named their system "Exposure Notifications" (EN) to emphasize its chief function from a smartphone user's perspective. The EN system was limited in scope. It was composed of a decentralized architecture for privately and anonymously logging a two-week history of close encounters with other EN users, and a centralized architecture for privately and anonymously alerting other EN users of a positive COVID-19 exposure. Apple and Google jointly devised technical specifications for the Bluetooth Low Energy (BLE) messages to be broadcast by smartphones, and for the construction of daily cryptographic Temporary Exposure Keys (TEKs) and rolling proximity identifiers (RPIs), which would be used in place of personally identifiable information (PII). [10] [11] Over the summer of 2020, individual U.S. states began to pilot and deploy smartphone apps using the EN service and application programming interface, in the absence of federal willingness to develop a nationally available app. To provide interoperability between the apps for individual U.S. states, a national server architecture was created through the support of CDC, the Association of Public Health Laboratories, and Microsoft. [12]

Figure 2 describes an example encounter between two users of any EN implementation, Alice and Bob. Each day their phones begin using fresh individual TEKs to generate a series of RPIs. **(A)** Throughout the next 24 hours, each phone emits Bluetooth messages, called "chirps", containing an RPI that changes approximately every 15 minutes, and some encrypted metadata, including the transmit power, which is hardware-dependent. EN keeps track of the messages they each receive for up to 14 days and the received signal strength indicator (RSSI) and timestamp of each message. **(B)** When Alice tests positive for COVID, she receives a one-time code from her public health department, at the phone number associated with her laboratory test. **(C)** If she clicks on the link or manually enters the code into the app, the app communicates with the national verification server and validates the one-time code, receiving a signing certificate. The signing certificate enables Alice to anonymously upload the last 14 days of her Exposure Keys to a national server. The keys are marked with metadata that help to estimate Alice's infectiousness on a given day. **(D)** All EN-running phones check the key server several times a day for new exposure keys to download. When Bob's phone downloads keys, it regenerates all of the RPIs for those TEKs, and reviews its history for matching RPIs. If there are matches, the timestamps and signal strengths of the matching messages are fed into a risk scoring algorithm devised by Apple and Google, which can be tuned by the PHA in its EN configuration. [13] **(E)** If the total score exceeds the PHA's predefined threshold, Bob sees an alert on his smartphone, advising him of the risky exposure and recommending next steps (e.g., testing and self-quarantine), alongside contact information for his local PHA.

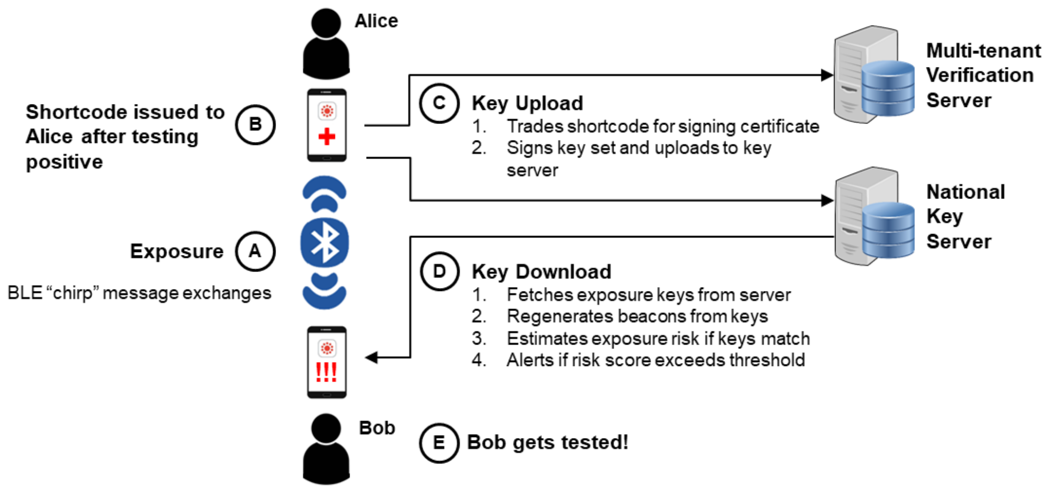


Figure 2. Exposure Notifications operational view.

The example in Figure 2 shows only two users, but the system was designed to handle multiple and overlapping exposures on each day. The EN risk scoring algorithm uses a set of weights and thresholds to assign a risk score to each exposure, and sums the score of each exposure to arrive at a daily cumulative score. Much of the work conducted under Program 10383 centered on the technical and practical implications of EN weights and thresholds, and is discussed in depth in the ACTA series of program reports. [14] [15] [16] [17]

The first U.S. states to launch EN apps did so during the summer of 2020, as Google and Apple revised the initial design of the risk scoring algorithm, and by extension, the application programming interface (API). Other states took a “wait and see” approach, or wished to deploy an EN app that would be simple to maintain and would not require state resources to keep up with future API changes. To lower the bar for adoption by PHAs, Apple and Google produced EN Express (ENX) in the fall of 2020. [18] ENX was a “turnkey” version of Exposure Notifications, which enabled jurisdictions to launch the service with their chosen branding, risk scoring configuration, and user-facing messages, without having to develop an application themselves or contract out the development to a vendor. Apple and Google worked with each state to prepare its ENX package for release, and the state managed its ENX settings, including messaging, through the Apple Business Registry website. On iOS, the ENX user interface was managed through the operating system settings (and updated through OS version updates). On Android, the ENX user interface was deployed as an app through Google Play Services. In April 2021, Google made it possible to deploy ENX as a “headless” app, so that it would look more like a built-in Settings module to the user, in hopes of increasing user adoption and reducing attrition. In 2021, several states that originally deployed a custom app (2020) also launched an ENX version in order to boost adoption; as of June 2022, some of those states were retiring their custom app and migrating all users to the ENX version, in order to simplify EN support within their states and/or use the ENX analytics system.

This page intentionally left blank.

4. SUPPORT TO U.S. STATES

MIT LL conducted two pilot projects with U.S. state-level officials in 2020 and 2021, intended to assess the efficacy of a GAEN-based solution customized for each state's needs, and provided technical advice to the states during and after the pilot periods. The information gathered during each pilot helped to inform the parallel research activities of the program, and each pilot partner state shared the lessons learned with other U.S. states and with system implementors. MIT LL and CDC also co-hosted a series of "Learning Lab" meetings with all interested U.S. state partners, which began concurrently with the first state pilot, and continued through June 2022. The Learning Lab sessions and the pilot studies formed the bulk of MIT LL's direct support to U.S. state public health practitioners under Program 10383 (Figure 3).

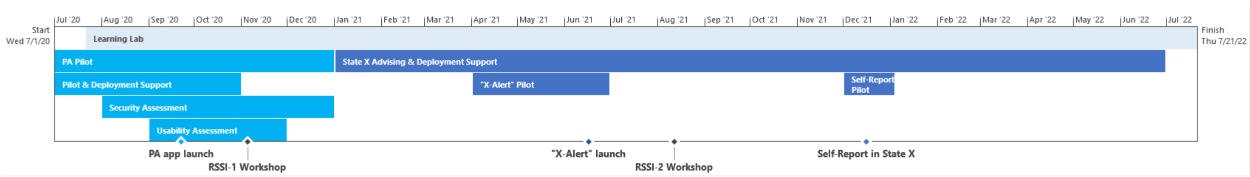


Figure 3. Timeline of support to U.S. PHAs.

Two shorter efforts also turned out to have lasting impact on U.S. states and on PHAs beyond U.S. borders. In November 2020, Linux Foundation Public Health (LFPH) convened an international workshop to come to consensus on appropriate "starter pack" configurations for PHA deployments of EN. MIT CSAIL and MIT LL staff presented work from Program 10383, which shaped the discussion and consensus definition. The workshop reconvened in August 2021 to reevaluate and update the template configurations, again with significant support from MIT LL staff.

4.1 COMMUNITY OF PRACTICE MEETINGS: LEARNING LAB

4.1.1 Purpose

The purpose of the Exposure Notifications Learning Labs was to assist public health agencies (PHAs) in implementing and enhancing the novel EN protocol for COVID-19 response. The Learning Labs started as an informal venue for PHAs to share knowledge and experience, collaboratively problem solve, and consult with technical experts. Over time, as participation grew to every state that implemented EN, the Learning Labs became a central hub for collecting and disseminating critical technical and practice information, which helped PHAs integrate EN for COVID-19 disease control and message delivery nationwide.

4.1.2 Methods

Learning Labs were designed to accelerate peer-to-peer learning by facilitating knowledge sharing and collaboration. Convened on a semi-monthly schedule, Learning Lab participants met via Zoom for a facilitated conversation with a semi-structured format. Agendas were organized in four sections: (1) welcome orientation and ground rules, (2) announcements, (3) learning sharing, and (4) the focus topic. The Chatham House Rule governed every meeting, creating a forum in which state officials and partners could speak openly about their successes and struggles using EN. [19] From one meeting to the next, sections varied in content and duration to meet participant interests and priorities, as EN features evolved in response to changing pandemic circumstances. Initial Learning Lab participation and invitations were handled by CDC, while scheduling and communications were handled by MIT LL. A facilitator experienced in convening public health professionals on national matters concerning emergency preparedness and informatics led the Learning Labs. Facilitation tools included live polling, mind mapping, and surveys.

Statistics on Learning Lab participation were calculated using Zoom attendance data. Anonymized Learning Lab meeting notes captured by the facilitator on mind maps were used to characterize discussion topics. Data from a March 2022 participant survey were used to capture the utility of Learning Labs and shared with CDC.

4.1.3 Results

Participation. Over 23 months (Aug 2020 to Jun 2022), thirty-two (32) Learning Lab meetings convened a group of participants that included a representative from every EN implementing state public health agency (Figure 4). The participant group included state health officials (55%), federal officials (14%), operational service and technical assistance providers (18%), and non-governmental organizations (5%).[3] Meeting attendance varied from 30 to a maximum of 57 professionals. Of the 29 states, 12 states (41%) attended 10-19 meetings and 7 states (24%) attended 20-32 meetings.

[3] 8% of participants were unidentifiable by affiliation in Zoom records, i.e., had called in from a cell phone or used a screen name that did not map to a known email address.

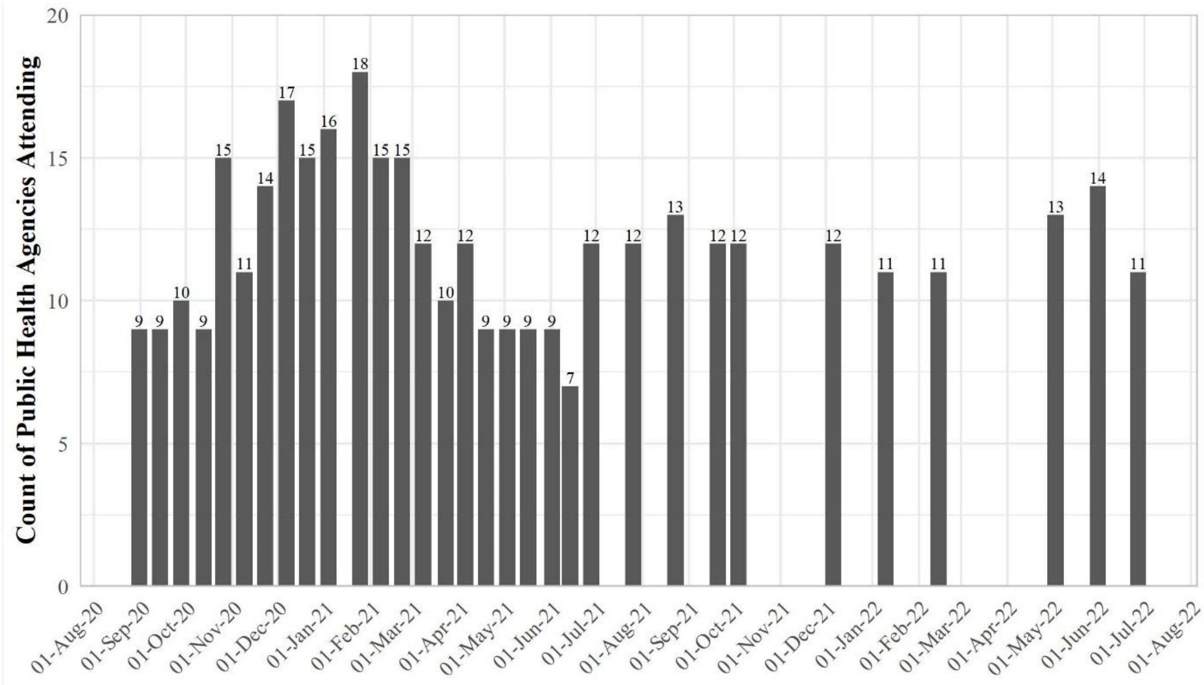


Figure 4. Learning Lab meeting attendance by representatives of U.S. public health agencies (Aug 2020 to Jun 2022).

Learning. Participants attended Learning Labs to learn how to assess, implement, and operate the novel EN technology. Over time, discussion topics changed as the technology advanced, experience in public health programmatic use grew, and communication challenges evolved (Table 1).

Table 1

Key Topics Discussed by Learning Lab Participants Regarding the Technical, Programmatic, and Communication Aspects of EN, Aug 2020 through Jun 2022

	Technical	**Programmatic**	**Communication**
Aug - Sept 2020	Verification server	Risk settings	Resource use
Oct - Dec 2020	EN Express	Workflow integration	Marketing
Jan - Mar 2021	Automated code distribution	Claim rates	Alert messaging
Apr - Jun 2021	SMS Intercept	Post-vaccination EN	Risk-based messaging
Jul - Sept 2021	Self-attestation function	Accommodating home testing	Back-to-school marketing
Oct - Dec 2021	ENPA metrics	Contact tracing capacity reduction	Multilingual messaging effectiveness
Jan – Mar 2022	Updated risk scoring for variants	Launching self-report feature	Twilio SMS best practices
Apr – Jun 2022	GAEN analytics tool	Self-report feature effectiveness	Reminders to self-report, changes to guidance

Table 2

EN Topics about Which Participants Shared Knowledge, Collaborated, or Heard about from Operational Service and Technical Assistance Providers during Learning Lab, Aug 2020 through Jun 2022

	Sharing knowledge	Collaborating to solve challenges	Receiving critical information from partners
Aug - Sept 2020	Marketing methods	Defining and measuring success	Implementation of National Key Server (NKS) and Multitenant Verification Server (MVS)
Oct - Dec 2020	Factors affecting implementation/adaption decisions	Risk settings with testing capacity	SimAEN
Jan - Mar 2021	Success stories	Risk setting with vaccine	ENPA implementation
Apr - Jun 2021	ENPA use	Measuring impact	NKS and MVS operations
Jul - Sept 2021	Effects of SMS intercept	Secondary attack rate estimation	Updated risk scoring recommendations
Oct - Dec 2021	Cross-national implementation	EN impacts Morbidity and mortality	Apple/Google update on metrics
Jan – Mar 2022	Experiences with self-report launch	Barriers for enabling self-report	RSSI-2 risk scoring workshop
Apr – Jun 2022	Trends in self-report use	Advocating for self-report feature	GAEN analytics tool demo

Utility. In March 2022, representatives from eight state public health agencies reported that the Learning Labs were very helpful to their EN work. Most valued was the ability that Learning Labs provided for states to share EN knowledge, collaborate to overcome challenges, and receive critical information from partners (Table 2).

4.1.4 Summary

For the adoption and implementation of EN during the COVID-19 crisis, the Learning Lab was essential to accelerating technological efficacy. It was a valuable capability that can be activated in service of public health challenges that exceed the scope of any single community, approach, or skill set.

4.2 PENNSYLVANIA PILOT AND TECHNICAL SUPPORT

Within the U.S., Pennsylvania was one of the first states to explore the adoption of EN. CDC instructed MIT LL to work closely with Pennsylvania's Department of Public Health and to provide technical guidance to the design, evaluation, and deployment considerations associated with their EN exploration. The team quickly focused its technical efforts on Pennsylvania's proposed EN system and provided guidance in the following areas:

- Technical risk analysis
- Protocol functionality, especially pertaining to preserving user privacy
- EN integration with existing public health processes and data flows
- Test and evaluation of deployed EN systems

MIT LL became an integrated member of Pennsylvania's EN launch team and worked with personnel from the PA Department of Public Health; Pennsylvania's commercial partner, NearForm, which was building the state's EN app; and University of Pennsylvania researchers on a daily basis. This effort led to the successful launch of Pennsylvania's EN app, "COVID Alert", on September 22, 2020. The app was embraced by the Pennsylvania government and the governor encouraged residents to download and use the app to limit the spread of COVID-19. [20]

4.2.1 Pennsylvania App Risk Analysis

MIT LL's initial work with Pennsylvania started with the development of a framework to measure the success of the proposed application and the analysis of major risks that could affect a "successful" deployment. Early in the program, discussion focused on how to ensure the app would be a success. With the many different professional perspectives involved, there were different definitions of success and how to achieve it. To help clarify the meaning of "success", MIT LL defined a framework, illustrated in Figure 5.

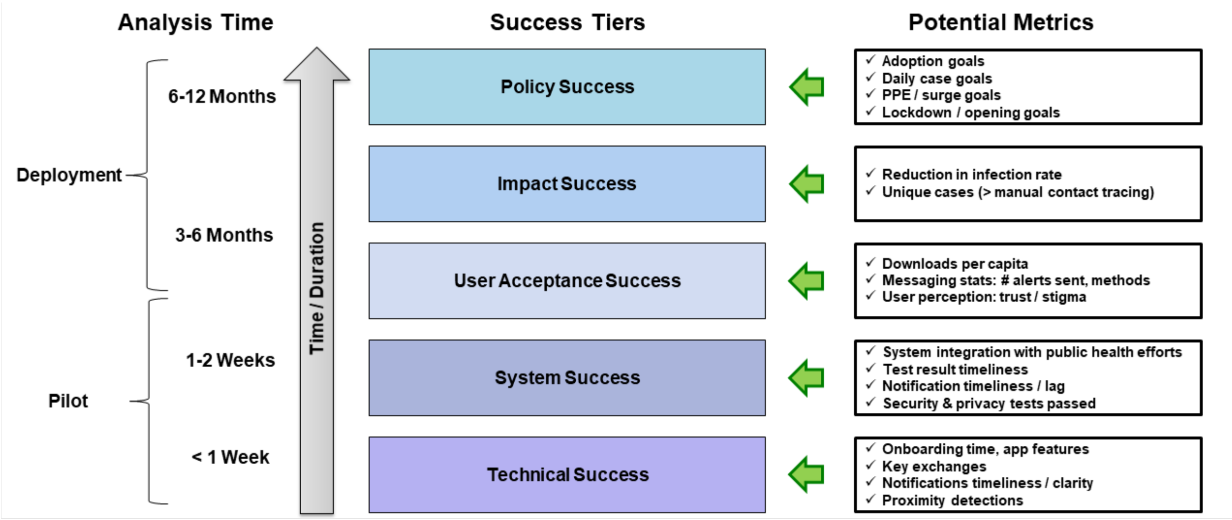


Figure 5. Framework for assessing EN deployment success.

The success framework articulated five tiers of success, ranging from technical success to policy success. It identified evaluation metrics for each tier, and estimated how long it would take to evaluate each level of success. This framework helped the various stakeholders discuss the levels of success important to them, and provided an indication for how success could be measured. For example, this framework showed that a one- to two-week evaluation pilot with the app would be able to provide evidence that the app was working correctly and that people knew how to use it, but it would not be able to determine whether the EN system was effective at reducing COVID-19 exposures and transmission rates.

MIT LL also helped Pennsylvania understand the technical risks associated with the deployment of an EN app. A summary of these risks and the associated risk weighting is shown in Figure 6.

	Risk	Consequence	Likelihood
1	Overwhelm the healthcare system	Severe	Possible
2	Privacy or security risks	Severe	Possible
3	Public health system delays of information transfer	Serious	Possible
4	GAEN detection performance poor	Serious	Probable
5	App not adopted by enough people	Serious	Likely
6	Alerts, instructions not followed	Acceptable	Unlikely
7	Version disparities	Minimal	Probable
8	App confusing/challenging to use	Significant	Possible

Likelihood	Minimal	Acceptable	Significant	Serious	Severe
Likely				5	
Probable	7			4	
Possible			8		1, 2, 3
Unlikely		6			
Remote					

Consequence

Figure 6. Summary of Pennsylvania's EN app deployment risks.

The risk matrix summarized the perceived risks associated with EN app deployment across various considerations, including healthcare systems, technical performance, and the social adoption of the application. As one of the first EN deployments in the U.S., Pennsylvania officials expressed concern regarding the technical performance of the app, i.e., whether the Bluetooth proximity detector was able to capture COVID-19 exposure events, and regarding the social adoption of the app. App adoption was particularly important because low adoption would directly impede the app's performance—i.e., if not enough people used the app, there would be very few EN key exchanges, which would cause EN to miss many real COVID-19 transmission events.

As part of the analysis of risks, MIT LL worked with Pennsylvania to address a large array of questions pertaining to the deployment of an EN system. The following is a sample of questions posed during this process:

- What are the performance goals (speed, detections, few false alerts, etc.)?
- What fraction of the population needs to use this for it to be effective?
- How do officials and citizens know the app and system are "working"?
- What risk scoring parameter settings are appropriate?
- Should these settings change for various geographic regions?

- How should alerted citizens respond?
- Will alerted citizens be referred to testing? Will this exceed current capacity?
- How do officials authorize and enable citizens to upload diagnosis keys?
- Which officials will have this authorization ability?
- How are diagnosis keys shared across jurisdictional boundaries?
- Is it possible to detect system abuse or misuse?
- Does the app leak private information?
- What public health information do you expect to gather from this system?

4.2.2 EN Protocol Functionality

MIT LL assisted Pennsylvania with an analysis of the GAEN protocol and its implementation within a custom mobile phone application developed by NearForm. Much of this analysis focused on understanding and explaining exactly how Apple and Google were implementing EN data flows, the associated privacy provisions, and the EN alerting alogorithms embedded in the Apple and Google implementations. MIT LL provided numerical assessments of the EN risk scoring algorithm and attempted to quantify the algorithm's parameter space (settings) and the associated parameter impacts on EN detection performance. MIT LL's privacy and security analysis and findings are covered in program report ACTA-1. [19]

4.2.3 EN integration with Public Health

The analysis effort led to the decision that a well-functioning EN system should be integrated with Pennsylvania's public health enterprise. To understand how this integration should occur, MIT LL worked with PA Department of Public Health to map the existing testing and contract tracing workflows and to identify integration points for the new EN technology.

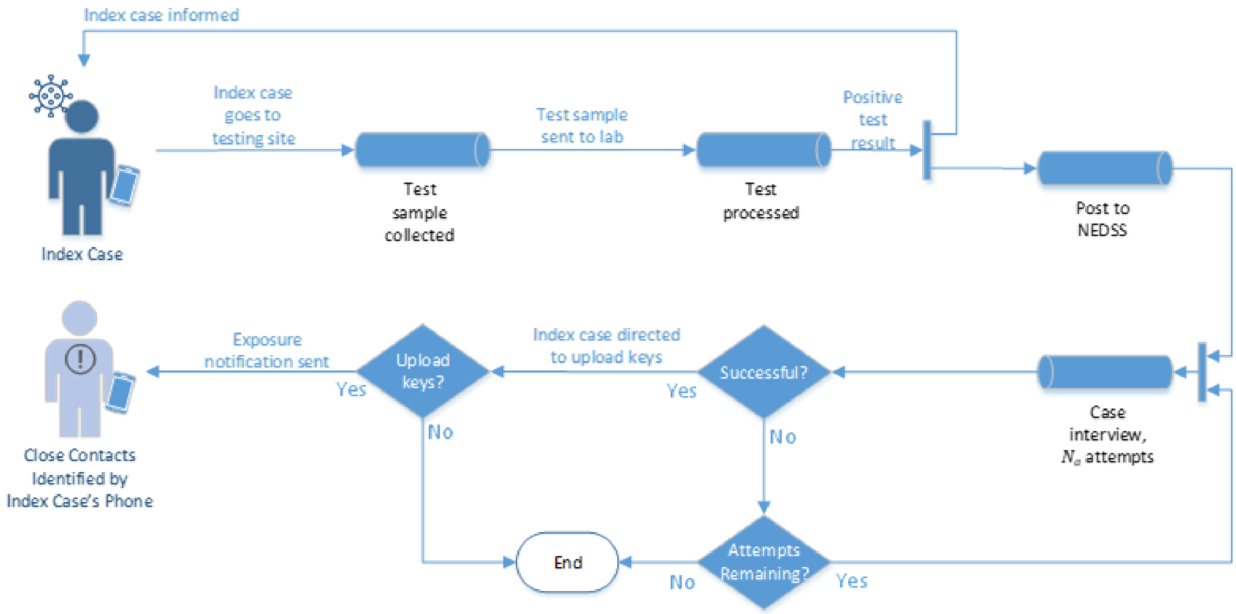


Figure 7. EN Integration with Pennsylvania's testing and tracing workflows.

Figure 7 illustrates, at a high level, how EN could be integrated with PA's testing and tracing workflows. The two primary integration points were with the case interview step and posting the case data to Pennsylvania's National Electronic Disease Surveillance System (NEDSS). The understanding of the workflows informed subsequent simulations for EN performance, particularly with MIT LL's simulated AEN system, "SimAEN."

4.2.4 EN Test and Evaluation

MIT LL also supported Pennsylvania's test and evaluation efforts for the EN capability, particularly with the NearForm instance of EN. MIT LL outlined a variety of pilot strategies for Pennsylvania to consider, ranging from small high-fidelity, well-controlled tests to large unconstrained evaluations, as shown in Figure 8.

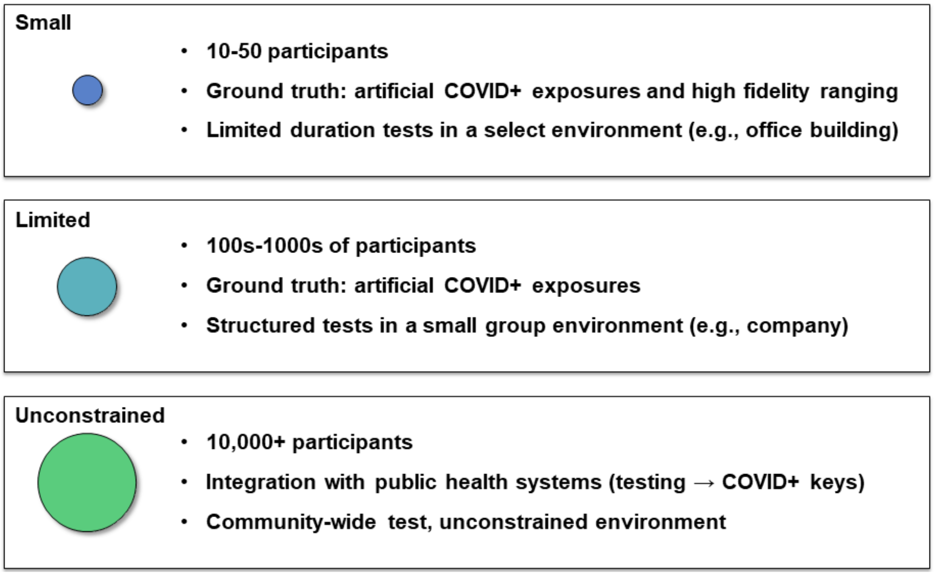


Figure 8. Notional evaluation pilots for PA's EN system.

Ultimately, MIT LL advocated for testing EN in three different environments: (1) full simulation, (2) laboratory test range, and (3) *in situ* pilots. A simulated environment would allow the team to explore the full parameter space of the EN application and experiment with different public health integration models. This would allow MIT to control almost every aspect of EN and test it in a large variety of hypothetical situations. The drawback with simulation was that it might not be an accurate representation of the actual EN and environmental physics, and outcomes would be highly dependent upon the modeling approaches. The laboratory testing would be an intermediate approach between simulation and *in situ* tests. The EN devices and protocol would be tested in real but highly controlled environments. This could approximate the real-world performance of EN, but would not account for the wide diversity of situations that would appear *in situ*. Finally, the *in situ* tests would offer a full evaluation of EN performance by performing tests in the wild. This would have high fidelity and value, but it would be challenging to interpret the measured performance results. The *in situ* tests would also require a large number of participants in order to obtain diversity across environments, confounders, and social interactions. The first two, simulation and laboratory testing, were pursued by MIT LL, but the third was not, because a large-scale, unconstrained test was exceedingly difficult to perform under the constraints of lock-downs and extreme social distancing. As a surrogate for an *in situ* test, Pennsylvania decided to offer a gradual rollout of their EN application with very close monitoring of its performance.

Prior to the structured rollout of EN, Pennsylvania tested the end-to-end workflows associated with the EN application. This consisted of deploying the application to a set of beta testers, some of which were given validation codes to artificially trigger an EN exposure alert. Pennsylvania evaluated the application's usability, basic functionality, controlled triggering of alerts, and public health follow-on activity. At the end of the evaluation period, MIT LL delivered evaluation guidance, EN technology assessments, advice on EN integration with public health systems, and EN usability assessments.

4.3 "STATE X" PILOTS AND TECHNICAL SUPPORT

In early 2021, MIT LL began advising a U.S. state's Department of Public Health (DPH) and staff from its contact tracing section[4] on the technical aspects of Exposure Notifications, specifically the EN Express (ENX) option.[5] The state wished to conduct a pilot of Exposure Notifications before deciding whether to deploy an ENX-based solution, referred to here as "X-Alert", within its purview. The objectives for the pilot study encompassed technical, operational, and user-centered concerns, summarized in Table 3. The state wished to be especially thorough in usability and functional testing, performing complete end-to-end integration tests with intentionally triggered exposures and alerts across both Android and iOS test devices, using the real-world key server and verification server. Because Exposure Notifications had already been deployed in neighboring states, and testing potentially could trigger alerts to unsuspecting citizens who might then quarantine, MIT LL was engaged to conduct controlled laboratory tests for the initial "Pilot 0" phase, which would validate operational readiness. The "Pilot 1" phase would be conducted with volunteer staff from two employers based in State X, and their family members, to scale up testing to approximately 100-150 devices out in the community. A third "dress rehearsal" phase would be run by State X DPH in partnership with one or two municipalities, before a go/no-go decision would be made for a statewide launch.

[4] The state has requested not to be identified, per the terms of its Non-Disclosure Agreement regarding pilot activities.

[5] For more details on ENX and its OS-specific implementation details, please see Section 3.

Table 3

X-Alert Pilot Objectives

Technical	• Is the State X configuration file implemented correctly (do technical performance and message content match requirements delivered to Apple/Google)? • Does the integration between State X DPH and APHL verification server correctly send verification codes to anyone with a positive test result?
Operational	• Does the state's operational design support the proposed EN process?
User support	• Are people able to effectively complete the ENX onboarding process on Android and iPhone with the available documentation and support? • Are people able to effectively upload verification codes to report positive cases on Android and iPhone with available documentation and support? • Is the help desk able to resolve the issues users have when using ENX? • Are the help desk training and available user support resources well calibrated to the most common issues users encounter? • Is the help desk staffed adequately for the anticipated volume of requests in a statewide launch?
User experience	• Is the activation process easy for users to complete successfully? • Is the positive case reporting process easy for users to complete successfully? • Are all in-app messages clear to users? • Communications effectiveness (publicity materials, user instructions, etc.)

4.3.1 Pilot 0: Laboratory Tests

MIT LL designed a test plan to exercise the proposed ENX instantiation for iOS and Android, under a controlled environment, and to identify whether the full system (phones and servers) performed as expected in each use case. This would give State X early confirmation that all the servers and verification codes were connected correctly. Our testbed consisted of:

- 6 Android phones (mix of Motorola, LG, and Google) running OS versions 8 through 11
- 3 iPhones (iPhone 7 and XR) running OS versions 13.7 and 14.3

MIT LL had obtained an "allowlisted" Google account that permitted the Android EN service to run in debug mode, logging the beacons sent and heard as well as the keys exchanged with the server. We lacked that capability with iPhones, so the data logged on iPhones was limited to manually logged records of when EN was activated, keys were exchanged, and alerts were received.

The Pilot 0 concept of operations is shown in Figure 9. For a simple end-to-end integration test, it was deemed sufficient to use the phones as stationary "tabletop" devices (as opposed to dynamic tests with mannequins and motion capture). In order to minimize the chances of accidentally triggering a false alert on "bystanders," we conducted our tests in spatial and temporal isolation from other phones that might have been running EN (late at night and in an isolated facility).

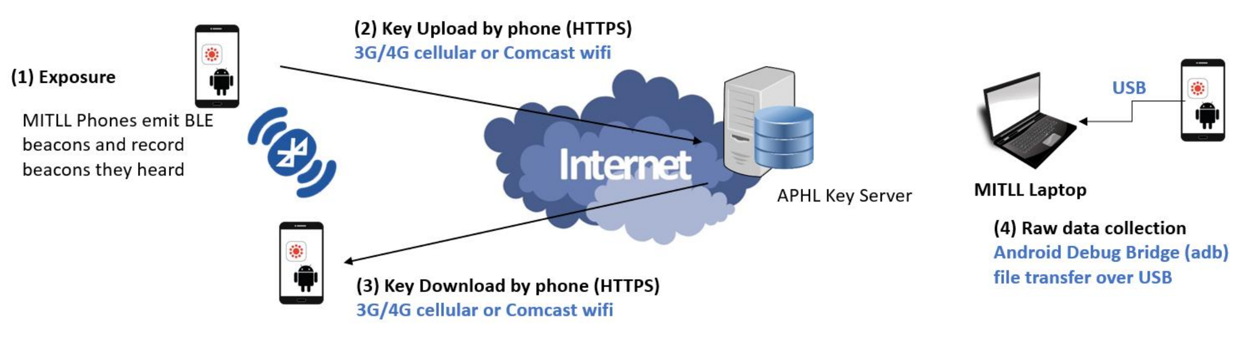


Figure 9. X-Alert Pilot 0 concept of operations.

The testing for Pilot 0 was conducted during two weeks in Spring 2021. Our detailed test sheet for Pilot 0 is included as Appendix A: State X Pilot 0 Test Sheet. MIT LL demonstrated that the candidate X-Alert implementation behaved mostly as expected on both Android and iOS, for modes of success (e.g., correct key upload) and modes of failure (e.g., attempting key upload when there is no network). MIT LL also identified a few modes of failure on iOS that did not behave as expected. In consultation with engineers at Apple, it was determined that one of these cases was due to repeat tests with a single phone in one UTC day, which we were told is not a code path a "normal" user can trigger; the test succeeded on a subsequent day. Apple later implemented a fix to the code path after it was shown in X-Alert Help Desk records to be the root cause of an incorrect error message for real-world users, as well.

The complete Pilot 0 test results are included as Appendix B: State X Pilot 0 Results. Based on the test results, State X DPH decided to proceed to Pilot 1.

4.3.2 Pilot 1: X-Alert Used by Non-Expert Personnel

The next phase of the X-Alert pilot involved recruiting a small population of adults from two large employers in State X, as well as any adult family members, to enable X-Alert on their personal cell phones and run it for two weeks. This phase of the pilot would not use contrived exposures with "fake" positives injected into the operational system; rather, it would rely on only "natural" positives among the participants to trigger exposure alerts. At the end of the pilot period, participants filled out a survey to send feedback to State X DPH about their experience using X-Alert. Figure 10 shows the pilot structure for one employer's staff (the experience was similar for the other employer's staff, and the enrollment website and survey were identical for both populations). Approximately 100 people participated in this phase of the pilot.

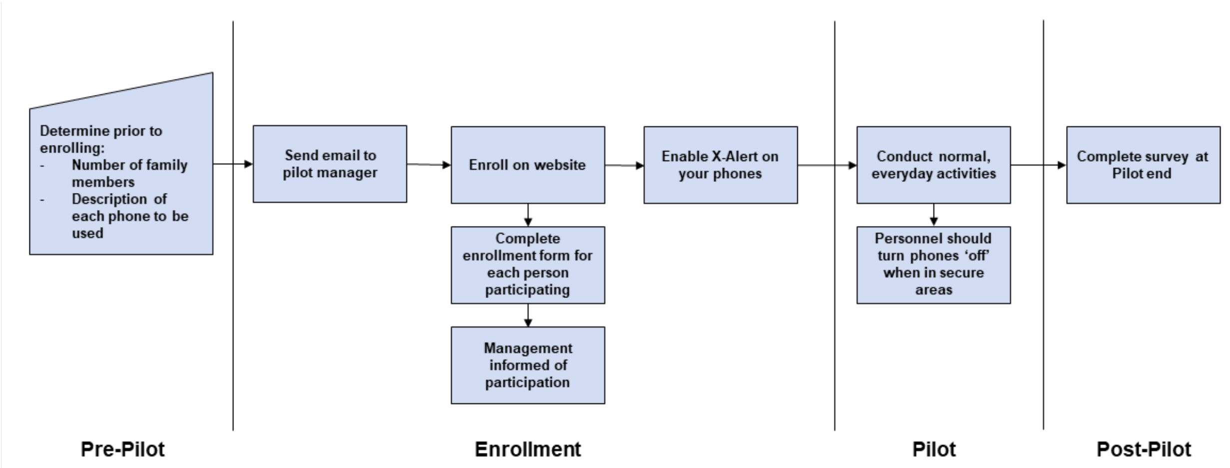


Figure 10. State X Pilot 1 flowchart for non-expert participants.

During the two-week test period, there were no natural positive cases among the pilot participants. Therefore, X-Alert support staff designed a readiness testing exercise for their Help Desk, in which MIT LL staff participated, to field calls from people who just received an exposure notification. As there were no red flags raised from the participant surveys or the State X DPH and Help Desk staff, State X proceeded to the "dress rehearsal" phase of its pilot in two municipalities. MIT LL assisted the X-Alert team with evaluating new features on Android when they became available during the dress rehearsal. Finally, X-Alert proceeded to a statewide launch in Summer 2021.

4.3.3 X-Alert Self-Report Pilot

In Winter 2021-2022, State X requested technical assistance with testing a new ENX feature, the ability of users to request a verification link within the app itself, rather than via a call to the help desk. This "self-request" or "self-report" feature had been piloted in Colorado and shown to increase key sharing by those who self-identified as having a positive COVID-19 test, while not permitting malicious repeated shares. Because an end-to-end test of the new feature would require enabling it in the already-deployed ENX interface, and injecting "fake" keys into the live service, State X DPH returned to MIT LL for test design and execution in partnership with the X-Alert support team.

MIT LL conducted a short series of end-to-end link requests and key-sharing tests under physical and temporal isolation, as in our original pilot. We established that the system worked as described, but identified a few key issues to be addressed (Table 4).

Table 4

X-Alert Self-Report Pilot Results

Issue	Correction or Mitigation
Differences in user screen messages from screenshots given to state team (Android)	• Update help desk support materials to match current version
SMS link sometimes not delivered to phone by Twilio service	• Investigated with APHL and Twilio support desk; told "delivery not guaranteed"; X-Alert help desk will prepare to support users who experience this
Expiration window for links was too short if link was requested ~9am ET, when state's batch notifications from X-IDS[6] went out to Twilio SMS Service	• Increase expiration window (Apple/Google) • Schedule X-IDS batch processing in smaller chunks throughout the day (State X DPH) • Establish separate SMS request queue on verification server (APHL)
Confusing iOS workflow in user interface	• Investigated with Apple and identified points of confusion for users • Traced one case back to incorrect error message identified in April 2020 Pilot 0 (Apple bugfix in next release) • Improved support desk understanding of workflow and language to assist users

The deployment of the self-report feature coincided with the onset of the Omicron surge in State X, the winter holiday period of increased travel, higher availability of home-use rapid antigen tests, and reduced capacity of diagnostic test clinics. The X-Alert team saw increased participation in EN key sharing and relied heavily on both the automated X-IDS link distribution, as well as the self-request feature, to empower State X citizens to notify each other of risky exposures.

In addition to the specific pilot efforts in partnership with State X DPH and contact tracing staff, MIT LL provided ongoing support in 2021 and 2022:

- Modeled the sensitivity and specificity of proposed EN risk scoring configurations, using predicted probabilities of alerting [16] and data collected in our experiments [14], to support State X DPH target sensitivity and specificity

[6] "X-IDS" is a pseudonym for State X's infectious disease surveillance database. It contains clinical, epidemiological, laboratory, and case management data in support of case investigation and disease surveillance for several dozen reportable infectious diseases.

- Reviewed user-facing messages and applying lessons learned from trust and user adoption research [22] to suggest ways to improve user experience and key sharing followthrough
- Improved understanding of help desk requests by reproducing “weird” cases in the lab and providing documentation to support bug reports
- Confirmed updates to user-facing screens, and recording sample workflows for help desk training
- Performed troubleshooting of SMS delivery issues with APHL, Apple, Google, and X-Alert staff
- Reviewed State X code- and key-claim data and ENPA metrics to check for anomalies and identify trends

4.4 RISK SCORING SYMPOSIA

In November 2020, the Linux Foundation Public Health hosted an invitational symposium to collaboratively develop risk score parameter guidance for health authorities who were using or planning to use the EN service in their regions. The invitees included Exposure Notifications implementers, practitioners, and analysts from industry, academia, and public health, including CDC and MIT as well as researchers and subject matter experts from the international EN community. The symposium met in the form of a two-day workshop, with presentations from two U.S. jurisdictions, EPFL, MIT LL, University of Maryland, and Google Research, and breakout sessions in which participants tackled the clinical, engineering, and modeling aspects of the EN risk score computation. The result of the workshop was a draft set of recommended templates for configuring EN, which was circulated throughout the EN community of practice for comment and refinement. The recommendations were finalized at the end of November, 2020. [23]

The recommendations of the workshop were intended as a starting point from which each PHA could make its own decisions, to meet their desired sensitivity/specificity “set point” at any given stage of the pandemic. For instance, in a region with high vaccination rates, the PHA might choose to use a less sensitive and more specific “narrow net” template (or a variation on it) so that only users with a very high “dose” estimate would receive alerts. Likewise, a region with lower vaccination rates and/or an extremely virulent variant of the disease might choose to use a more sensitive “wide net” template (or a variation on it), in order to communicate an appropriate level of risk and follow-on actions to users with even a low or moderate exposure.

The components of the templates included both the weights-and-thresholds settings for attenuation (determinants of “near” vs. “far” exposure impact) and the mapping of days-since-symptom-onset to presumed infectiousness. The structure and usage of these settings was predetermined by the API, that is, by Apple and Google; specific values for these data structures were proposed and refined by consensus of the workshop participants. Separate definitions for the attenuation and infectiousness parameters were developed for each of the narrow and wide configurations. During the November 2020 symposium, the overall risk score threshold was assumed to be 15 weighted minutes, in keeping with CDC's contact tracing guidance [9], and the group did not evaluate the impact of changing the risk score threshold.[7]

The narrow and wide templates, or close cousins of them, were adopted by most U.S. jurisdictions by the end of 2020 and into 2021. At that time, each jurisdiction had defined a single configuration and a single set of messages to users, although the EN service could support up to four classifications of alert per jurisdiction. The spread of the more infectious Delta variant in 2021 caused many EN-using jurisdictions to reconsider the applicability of their initial settings, and in August 2021, the Risk Scoring Symposium was reconvened to examine the utility of the templates in light of new epidemiological evidence and the progress made in BLE data collection and EN analysis since the first symposium.[8] MIT LL again contributed technical information[9] to the meeting, in collaboration with subject matter experts at Apple, Google, the University of Oxford, the Alan Turing Institute, the Robert Koch Institute, and U.S. jurisdictions.

The participants in the second symposium agreed that the original "narrow net" settings were too conservative. Data available from the new Exposure Notifications Private Analytics (ENPA) system showed "substantially fewer" notifications triggered than would be effective. Therefore, the new recommendations included an update to the "narrow net" weights and thresholds. [24] The infectiousness mapping was also updated, in order to reflect the higher infectiousness of the Delta variant, and for simplicity's sake it was recommended to use a single infectiousness curve regardless of the attenuation weights-and-thresholds chosen.

The participants in the second symposium also considered the use of different exposure risk thresholds within a single implementation, which was not discussed during the first workshop. The potential benefits of such an approach were twofold: first, to convey different levels of risk and follow-on actions to users; second, to provide more visibility for PHAs into the levels of exposure encountered by the users in their jurisdictions (ENPA provides aggregate metrics on the number of exposures for each classification). A PHA could choose to use different messaging for different levels of risk, or the same messaging for all.

---

[7] MIT LL's subsequent work on risk scoring analysis included a consideration of threshold changes (see Section 5.4).

[8] Section 5 describes MIT LL's work on BLE data collections and risk score analysis. Section 8.2 describes our analyses of the ENPA data available at the time.

[9] Results presented were derived from the BLEMUR model; see Section 6.2 and [16].

Example messages and tiered thresholds were provided as part of the final recommendations, based on the experience of Germany's system operators.

The two Risk Scoring Symposia provided the international EN community of practice with a forum in which to discuss and develop operationally relevant plans, and learn from each others' deployments as well as from CDC-sponsored research of MIT LL and the concurrent work done at academic centers. The version 2.0 configurations and tiered notifications continued to be the focus of discussion at the CDC-MIT hosted Learning Lab sessions, as the even more infectious Omicron variant displaced Delta, and as state-level contact tracing efforts were reduced or discontinued. Although a comprehensive analysis of deployments' configurations and ENPA metrics was not undertaken during Program 10383's period of performance, the desire of U.S. states to perform such analysis was there and the groundwork was laid for collaboration between PHAs, interested researchers, and system operators to achieve a more comprehensive understanding of the ways in which different configurations were deployed during 2022.

This page intentionally left blank.

5. DATA COLLECTIONS AND RISK SCORING ANALYSES

The collection of Exposure Notifications data on smartphone devices, and efforts to analyze the utility of both the Bluetooth Low Energy (BLE) detector and the risk scoring algorithm, were areas of significant effort on Program 10383 (Figure 11).

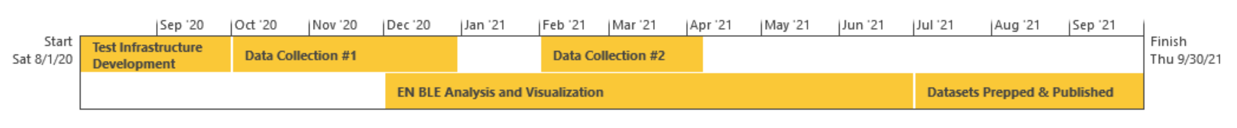


Figure 11. Timeline of data collection and analysis efforts.

5.1 DATA COLLECTION WITH CDC GUARDIAN APP

Initially, MIT LL planned to exercise one of the NearForm app implementations and test the "v1" risk scoring algorithm, using iPhones and Android devices, and recording ground truth distance and duration using the motion capture system in the Autonomous Systems Development Facility (ASDF) at MIT LL. [25] To maintain compliance with strict COVID-19 safety protocols in place at the time, the testing would use mannequins with similar RF absorption characteristics to humans, mounted on robotic platforms (Figure 12). By the time data collection commenced at the end of August 2020, the test protocol relied on a CDC-developed reference implementation of an EN app named "Guardian" and Apple and Google had released a new "v2" version of the risk scoring algorithm and application programming interface. Without access to a "v2" EN app, the data collection proceeded with a focus on collecting low-level EN data about exposures from the Guardian "v1" app, and replaying the exposure data on a desktop-based implementation of Google's "v2" reference risk scoring implementation. [26] The details and results of the first, Guardian-based data collection are discussed in an associated program report, ACTA-2. [14]

A notable limitation of this data collection is that it included only data from Android phones; Google had provided the developer-level access to EN system logs that enabled the recording of EN Bluetooth messages with timestamps, whereas Apple had not provided any low-level data access. Another limitation is that it included a single model of Android devices with "good" reference power calibration, in order to focus the analysis on the RF effects of the phones' placement and orientation to its environment. Nevertheless, it provided a crucial, independent first look at the performance of the Android implementation, and laid the foundation for future experiments.

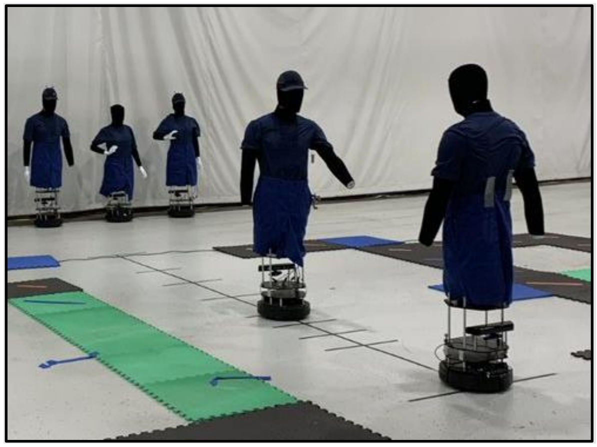


Figure 12. Mannequins with black RF-absorptive foam and phones placed in hip pockets, in ASDF test space.

Figure 13. Mannequins with phones in shirt pockets, aboard light transit railcar.

5.2 DATA COLLECTION WITH UNIVERSITY OF ARIZONA APP

MIT LL conducted a second data collection, in partnership with the University of Arizona's Campus Health Service, in February-March 2021. The University had deployed an Android- and iOS-based EN app to the campus population, and had conducted preliminary benchtop testing for app functionality. They wished to obtain an independent assessment of the app's performance in more realistic situations, using the same configuration of attenuation weights and thresholds, and two levels of risk scores, already in use on campus. CDC funded MIT LL to conduct the data collection and analysis with the objective of enriching the BLE exposure dataset, using various models of phones on both operating systems. The MIT LL team drew upon their experience with the Guardian data collection to develop a test plan for nine unique scenarios across social, professional, and public transit environments (Figure 13), and extended the software tools developed for the first data collection. The test scenarios included scripted "natural" mobility, up to 16 phones per scenario, and were repeated three times each. The team's earlier experience with Android EN data collection techniques enabled them to carefully construct test scenarios that included iPhones, and reconstruct in post-analysis which BLE messages came from which iPhones. The details, results, and practice implications of the second data collection are discussed in the associated program report, ACTA-3. [15]

5.3 DATA PRODUCTS

The data products of both data collections are available as a PACT github repository. [27] A Python visualization tool for the scenarios in the second data collection is also available, as the ability to replay scenarios and view the attenuations in relation to distance between phone pairs became extremely helpful for both the MIT LL team and the University of Arizona team. [28]

MIT LL presented technical summaries of the attenuation data to multiple states, agencies, and research teams during 2021-2022, via the Learning Lab sessions, the Risk Scoring Symposia, and in virtual meetings with EN partners in the U.S. and abroad. Two early users of the dataset were subject matter experts at Apple and the NHS/Oxford/Turing team. Internal to MIT LL, the data were used to inform the modeling work that resulted in BLEMUR and SimAEN (Section 6).

5.4 RISK SCORING ANALYSES

The risk scoring analysis efforts aimed to answer two questions, given current understanding of COVID-19 transmission: first, was the scheme proposed and implemented by Apple and Google a "good enough" approximation of a dosage estimator; second, how should the risk score configuration be used by public health authorities to achieve desired system characteristics?

In July and August 2020, MIT LL performed a mathematical analysis of the parameter selection space, in order to understand how to select values for each parameter of the exposure risk score framework. This analysis focused exclusively on the "v1" risk scoring algorithm, and ascertained that the initial parameter space was far larger than necessary to represent the physical ranges of interest. Concurrently, Apple and Google revised their algorithm to simplify the attenuation and infectiousness parameter space, and MIT LL refocused efforts on the data collections to support analysis of the "v2" implementation. The general architecture of those analyses is shown in Figure 14, and the methodologies and results are described in detail in the associated program reports. [14] [15] In brief, the data and analyses showed that the risk scoring algorithm is sensitive to the same effects of obscuration (RF absorption) and multipath effects (RF reflection). However, the analyses also confirmed that there is a subset of "useful" configurations, i.e., when examined across aggregated exposures, encounters that are "risky" due to higher duration/proximity are far more likely to trigger an alarm, and encounters that are "lower risk" due to brevity or greater distance are far less likely to trigger an alarm.

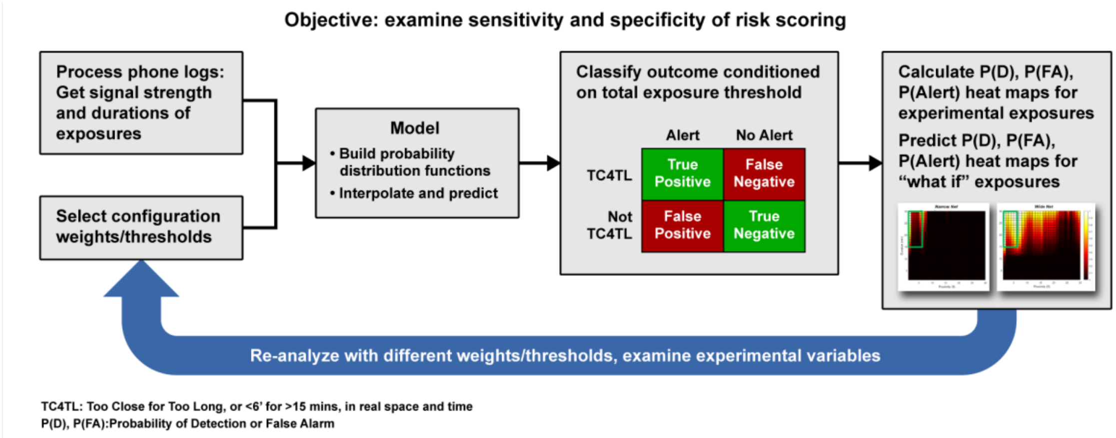


Figure 14. General approach to risk score analysis, relying on BLE data from Exposure Notifications.

5.5 SUGGESTED FUTURE WORK

While the data collections under Program 10383 were extensive, they were by no means exhaustive. These efforts uncovered several desirable improvements that could be made by the EN system implementers on hardware that is already available:

- More rigorous calibration of BLE transmit and receive power, preferably by the original equipment manufacturers as a prerequisite for device certification
- Revision of the range of appropriate BLE transmit and receive power levels, to more accurately represent the range of detectable signals above the noise floor on a variety of hardware models
- Reconsideration of whether the transmit power really must be part of the encrypted metadata (sending it in the clear could enable a consensus-based method of distance estimate refinement)
- Developer access to iOS-based system logs of EN BLE records, to enable broader and faster independent verification of the iOS-based implementation of EN

Bluetooth as implemented in 2020-2022 was only a coarse estimator of distance. The data collection and analysis efforts suggested a few approaches for future development of an EN-like system with improved distance estimation:

- Incorporation of new sensor hardware (e.g., ultra-wideband distance estimation triggered by BLE detection)

- Incorporation of additional BLE hardware (e.g., devices with two Bluetooth antennas could record angle-of-arrival and angle-of-departure) in combination with peer-to-peer message "echoing" to enable a consensus-based method of distance estimate refinement

Increasing the size of messages sent over-the-air, the amount of data necessary to store for up to 14 days, the computational complexity, and the number or type of antennas in use, would each come at a cost of storage and/or battery consumption. Future research in this area would help to define upper and lower bounds on those tradeoffs, and enable more informed design decisions.

The risk scoring algorithm provided several ways in which an individual user's risk score could be made more accurate, or which could provide the user with the context they need to interpret their exposure risk:

- Implement a more fine-grained permissions model, by which the infected user could permit the exposed user to see the approximate time of their encounter; the exposed user could then use their extra knowledge (I was alone, so it must have been through a window or wall; I was masked; I was outside; etc.) to gauge their true risk

- Incorporate a method of sensing whether the user is indoors or outdoors at the time of the exposure (e.g., changes in GPS signal strength trending with changes in cellular network signal strength)

- Incorporate a method of sensing whether the user is in a space with adequate ventilation at the time of the exposure (e.g., onboard CO_2 sensing as a proxy for air changes)

- Incorporate a method of logging whether the user is masked (e.g., using vocal biomarkers and/or visual mask detection)

None of these proposed ideas is a silver bullet, but each may have some merit, and several already are being researched for commercial purposes. Combinations of some of these approaches, particularly if augmented with machine learning methods and a rich experimental dataset on phone usage, have the potential to make an EN-like system capable of more useful risk estimates for COVID-19 or other airborne and/or aerosolized diseases.

This page intentionally left blank.

6. MODELING EFFORTS

6.1 OVERVIEW

The smartphone market is large and diverse; in 2020 alone, around 1300 new Android models went to market from 26 distinct vendors, and Apple released five new models of the iPhone. In December 2020, Google's BLE calibration spreadsheet included over 500 unique models that had been submitted to the RSSI calibration procedure over a range of antenna orientations and under controlled conditions, and therefore a tester or developer could be reasonably confident that EN should work as intended on that model. [29] This was an unmanageably large set of devices, should one wish to perform comprehensive testing or data collection *in situ* on the smartphones. Therefore, MIT LL pursued the development of models in order to make predictions about the performance of the EN Bluetooth detector and the risk scoring algorithm before the pandemic was over (Figure 15).

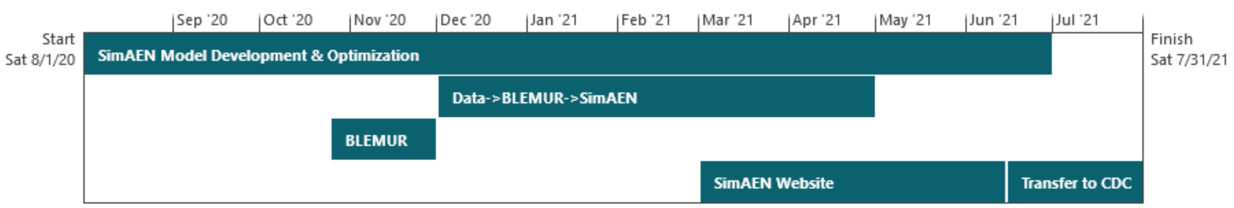


Figure 15. Timeline of modeling and simulation efforts.

6.2 BLEMUR AND HEAT MAPS

Our exploration of early BLE datasets and experimentation with mathematical models of the EN risk scoring algorithm led to the development of the Bluetooth Low Energy User Risk (BLEMUR) model from November 2020 through January 2021. For a given EN configuration of weights and thresholds, BLEMUR provides predictions about the probability of alert. If BLEMUR is also provided a list of potential "close contact" encounters, represented as distance-and-duration pairs, it will generate predictions about the probability of detection, probability of false alarm, and false discovery rate. The model mechanics are detailed in program report ACTA-4. [16]

To provide a useful visualization of which distance-and-duration contacts are most likely to be alerted—and help public health teams consider the implications of using one EN configuration versus another—BLEMUR used a contact "grid" representing all distance-and-duration contacts over a range of 1-30 feet and 1-30 minutes, and produced a probability of alert (*P(Alert)*) heat map. A notional example of this heat map is shown in Figure 16. CDC's definition of a close contact for COVID-19 is outlined in the upper left region of the contact grid, as a reference region. [1] A "perfect" detector would have a *P(Alert)* of 100% for every cell in the reference region, and a *P(Alert)* of 0% for every cell outside the reference region. The example shown is not perfect, but in public health terms, it has 100% *sensitivity*—alerting all "true" close contacts per CDC's definition—and moderate *specificity*—alerting many contacts outside the reference region.

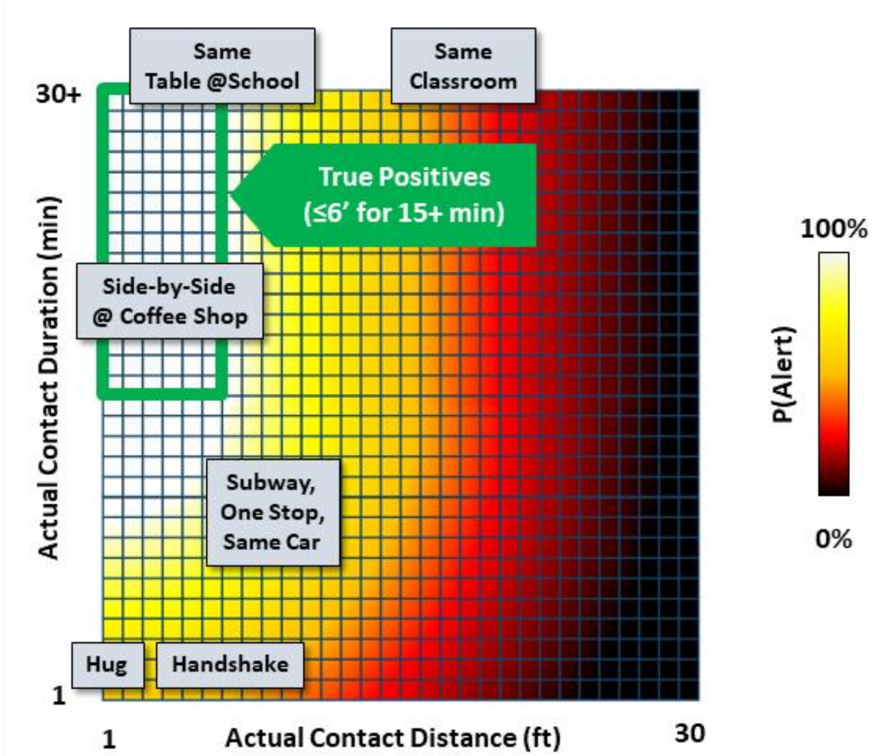


Figure 16. Notional P(Alert) heat map over uniform contact grid.

The heat maps produced by BLEMUR had an immediate impact within the EN community of practice, as a visually rich and easy-to-understand snapshot of how a given EN configuration might perform across a variety of encounters. One U.S. state's Department of Public Health[10] requested to view heat maps for variations on the "narrow net" and "wide net" templates before selecting a configuration for their EN launch, and reported that the heat maps were a deciding factor in their leadership's approval of the launch

[10] Section 4.3 describes MIT LL's support of this state's pilot, launch, and follow-on analysis.

plans. The work was presented to the U.S. EN community of practice meetings[11] as well as to the participants of the Risk Scoring Symposia[12].

While BLEMUR made good use of the BLE data collected at MIT LL and helped jurisdictions decide how to configure EN, it also highlighted some shortcomings of both the BLE-based EN risk scoring (sensitivity and specificity are inversely related, and EN would need "side" information about encounters to increase specificity alone) and of distance-and-duration (even if perfectly known) as a proxy for the actual risk of COVID-19 infection. The rapid evolution of COVID-19 variants, the uneven distribution and uptake of vaccination, variation in individual susceptibility, and the vagaries of airborne pathogen loads at any point in space and time all complicate the problem of estimating actual risk. Nevertheless, BLEMUR showed that approximations of risk and probabilistic predictions of system behavior are still useful to public health decision makers and system designers.

6.3 SIMAEN AGENT-BASED MODEL

If we consider a perfect smartphone-based detector of COVID-19 dosage and infection risk, with no false alarms and no missed detections, that system would still have to operate in a larger pandemic context. Because EN was architecturally a type of distributed sensor network, and because COVID-19 spread through social interactions, the network-level effects of both individual behaviors and public health decisions could affect the system-level effectiveness of an EN deployment. People may choose to share their EN keys and alert close contacts, or they may delay before doing so; they may choose to follow quarantine and isolation instructions carefully, loosely, or not at all. Public health authorities needed to consider factors such as lab and home test availability and reliability, processing timelines, new variants, and the easing of social restrictions and mask mandates. Because EN was designed to be anonymous and private and used no contextual or personally identifying information, it was almost impossible to obtain high quality population-level data on EN effectiveness.

To assist public health decision makers in understanding potential system-level effects, MIT LL developed a simulation model that incorporated relevant aspects of public health workflows, personal behaviors, the current understanding of COVID-19 transmission, and the PHA- and EN-specific interactions that would affect public health outcomes (e.g., whether people use EN, whether they answer a call from contact tracing staff, whether they decide to alert close contacts through EN). The simulation would need to predict outcomes such as COVID-19 case counts and transmission rates, the number of calls to and from public health staff, the number of tests used, and the number of people in quarantine or isolation. Using a simulation provides one advantage over real-world disease surveillance data, in that it would also be able to predict estimates of the number of people *incorrectly* quarantined/isolated (due to not actually

---

[11] Section 4.1 describes the "Learning Lab" sessions.

[12] Section 4.4 describes the Risk Scoring Symposia.

being a close contact, or not infected but tested positive) and the number of infected people unidentified by public health (due to being asymptomatic or unreachable). The former was an important cost metric, and the latter helped to illuminate the potential benefit of EN as augmenting traditional, labor-intensive contact tracing. Finally, simulations can help to predict bounds on common-sense relationships and feedback effects in complex systems: we all understood that reaching close contacts *faster* should help flatten the curve faster, but how fast is fast enough in a particular context?

To help address these questions, MIT LL began working on an agent-based model in August 2020, dubbed "Simulation of Automated Exposure Notification" (SimAEN). Agent-based models are well suited to predicting messy, human-centered situations. Each "agent" in SimAEN represented a person, and the simulation advanced day by day. The world in which the agents live was governed by rules, in this case, our understanding of the different phases of COVID infection and recovery, and the probability that an infected person would spread the disease. Users could also choose probabilities for mask wearing, social interaction levels, and compliance with quarantine, and could set levels for the number of contact tracers, the test turnaround time, and so forth. The simulations predicted relevant outcomes, such as the number of cases, number of calls to and from public health, and number of quarantines. MIT LL's early work with the Pennsylvania Department of Health (Section 4.2) and the guidance of CDC representatives were instrumental in developing the contact tracing and case investigation workflows within SimAEN. We validated the model using documented EN experiences from the UK and Switzerland; full implementation and validation details are described in program report ACTA-5. [17]

Notably, the model of EN within SimAEN abstracted away the implementation details of EN itself. The interaction model closely followed that of the Google-Apple EN implementation—in the choices available to individuals and to public health decision makers—but the detector and risk scoring were represented simply by parameters for the probability of detection (*P(D)*) and the false detection rate (*FDR*). This also eliminated the need to computationally represent individual encounters' distance and duration, reducing the complexity and runtime of the simulation. For the *P(D)* and *FDR*, SimAEN relied on BLEMUR predictions, which in turn were informed by the data collections MIT LL performed from 2020-2021. SimAEN was implemented in Python and relied heavily on the MIT LL Supercomputing Center for execution of over 1M individual simulation runs during development and testing. [30]

MIT LL presented initial findings from SimAEN to CDC in October to December 2020 and demonstrated a simple, MATLAB-based interface that would allow a user to vary the input parameters and compare the disease and resource metrics produced by the model. The initial results were well received and led to further iterations and refinements of the model's workflows and parameter space. However, the initial interface was suitable for use only by experts. CDC tasked MIT LL with parallel development of a web-based interface for SimAEN, which would be tailored to use terminology and concepts familiar to public health staff and decision makers, and which would need to produce results quickly without reliance on a supercomputer. The model implementation team worked closely with user experience subject matter experts and CDC advisors to produce a web app version of the tool. A screenshot of the interactive tool is shown in Figure 17, and the process used and value added by the user experience team are described in program report ACTA-6. [31] A CDC-MIT LL coauthored article in *Public Health Reports* shared examples of the type of investigations public health teams might conduct with scenarios in SimAEN. [32]

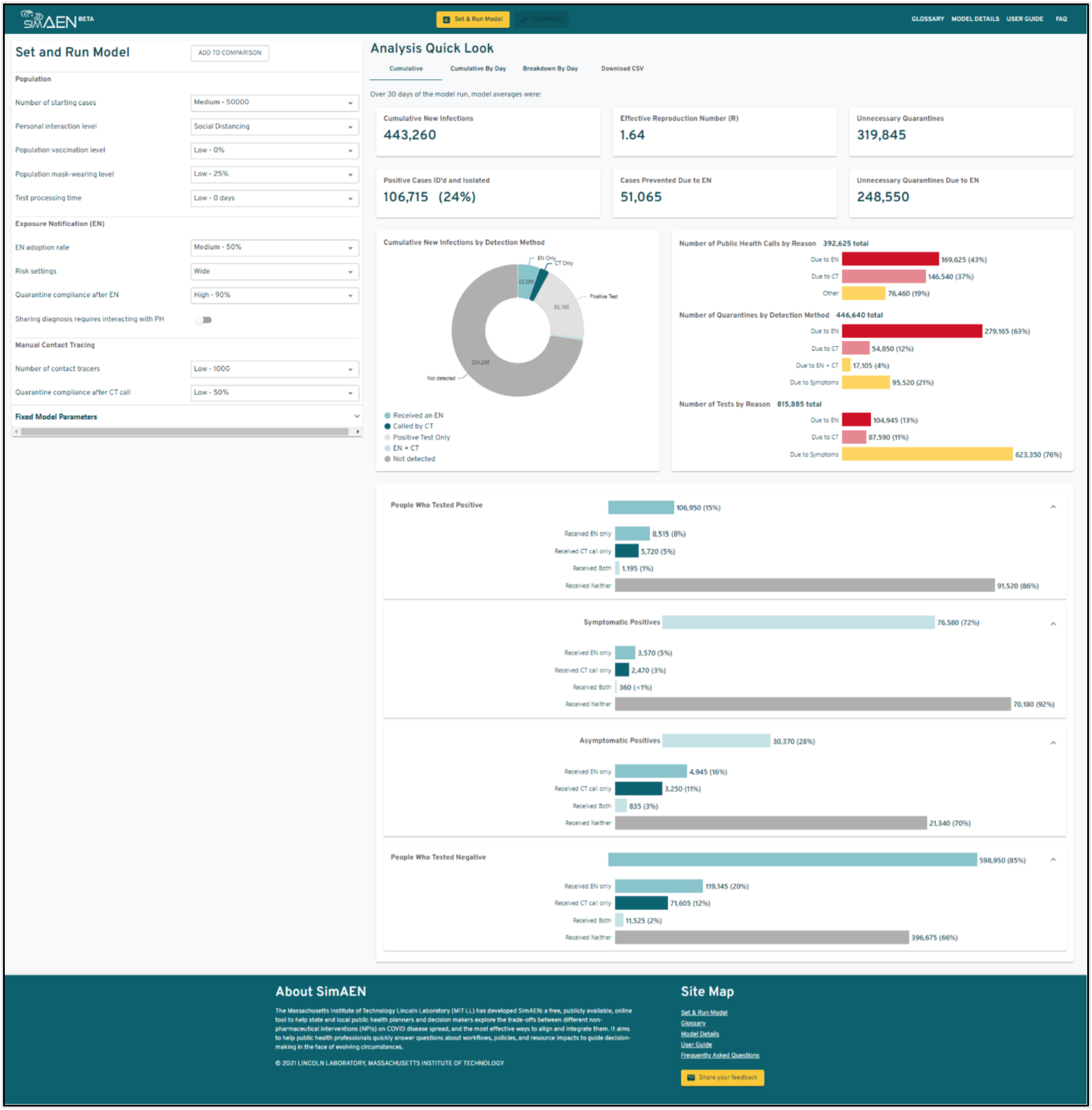


Figure 17. SimAEN model results for a single scenario, at https://simaen.philab.cdc.gov/.

The SimAEN web app was transitioned to CDC-hosted servers in July 2021, and the code for the model is hosted in a CDC Public Health Informatics Research Lab GitLab repository. [33] [34] To save computation time and serve results quickly, MIT LL and CDC identified “preset” parameter values that would be most relevant to users, ran a parameter sweep over the preset values, and stored the results. To smooth out stochastic effects, each set of initial values was simulated 30 times, and the results were aggregated. The aggregated metrics were stored in a database, which serves the website’s visualizations of simulation results. Users can compare model outputs from up to two runs, or download model parameters and outputs to produce more complex visualizations using their preferred software. The tool enables public health professionals and decision makers to explore critical transition points where interventions become effective.

6.4 SUGGESTED FUTURE WORK

Both the BLEMUR and SimAEN models would benefit by incorporating new data and knowledge about their constituent parts:

- BLE attenuation data with ground truth distance and duration for more hardware models and environments
- Human behavior data, such as how close people tend to stand and how many interactions they have of different types, which may vary across cultures and demographics
- Conditioning agent behaviors on circumstances (“if this, then that”), to account for more conservative social encounters and/or better compliance with recommended NPIs during intervals with more severe disease in the community
- Updating model parameters to account for more infectious variants, shorter or longer latent periods, etc. (these values are static in the web app implementation, but could be varied without redesigning the model)
- Updating model assumptions that were appropriate early in the pandemic, but no longer obtain (e.g., the assumption that a person only encounters one infected person in 14 days)

However, BLEMUR and SimAEN also suggest new areas of research beyond their application to COVID-19 and EN. SimAEN in particular is agnostic about the mode of disease transmission (airborne, fomite, intercourse, etc.) and the mode of “close contact” detection. This makes it especially adaptable to other diseases for which a suitable “infection risk detector” could be developed, now or in the future. In the final sessions of the Learning Lab meetings, several states expressed interest in investigating how EN might be adapted to help achieve better public health outcomes for other diseases, such as measles and avian flu. SimAEN or its next-generation version could help to answer key questions about how a notional system might perform under a range of conditions.

7. HUMAN FACTORS AND USER ADOPTION STUDIES

Despite the best intentions of technologists, the first generation of a novel technology is often imperfectly or even poorly matched to the assumptions, workflows, and cognitive needs of its intended users. Technologies that attempt to serve more than one type of user can fail to wholly meet the needs of either, without careful attention to the human factors (HF) and user experience (UX) aspects of design and implementation. Automated exposure notification systems, among other novel digital tools for coping with COVID-19, had the additional handicap of pandemic time pressure. For these reasons, MIT LL conducted research into the HF and UX elements of Exposure Notifications (EN) and similar systems, to help understand how users interacted with the technology, what they understood or believed about its functionality, and how that influenced their decisions to engage with EN and ultimately, use it (or not) to notify others of COVID-19 exposure (Figure 18).

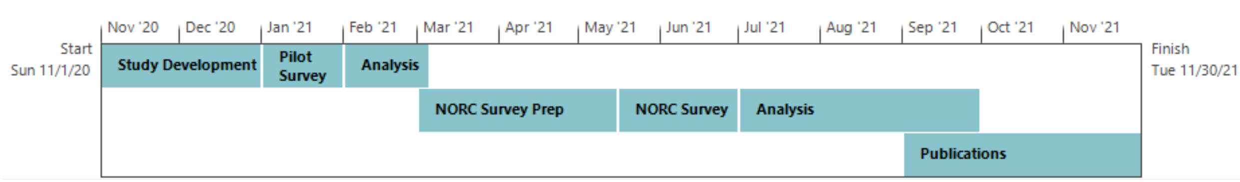


Figure 18. Timeline of trust and user adoption research efforts.

Initial work focused on the discrete user actions required to enable intended technology functions. Because EN only works when *both* persons in a close encounter have opted into the service ("adoption"), *and* the infectious person chooses to share their keys and notify others ("engagement"), there is some critical threshold for user adoption and engagement within a community below which EN will be ineffective. In short, if the infected people you encounter are not running EN or choose not to share their status, EN would never notify you of the infection risk.

The EN API mandated specific user opt-in actions before EN functions could be used: consent to the region's legal and privacy terms, activation of the EN service, entry of key metadata such as the symptom onset date, and consent to share one's keys and notify others of close contact. Each of these affirmative steps was presented separately; for instance, users could not pre-consent to key sharing when they chose to enable the service. In addition, smartphone screens are small and smartphone user interfaces rely heavily on multistep workflows to present bite-size amounts of information sequentially; each time the user has to tap a button to reach the next screen, they are choosing whether to continue with EN or to switch to one of the other myriad demands on their attention. Finally, EN services were rolled out in the midst of a pandemic, i.e., a time of heightened stress in much of the intended user community. Collecting information from users about *how* they engage with EN, and *why* they make certain choices, would help EN implementors as well as PHAs understand the impact of their design choices and the information they present to users at each step of the onboarding and key sharing workflows. MIT LL's objective was to obtain a richer understanding of users' needs, perceptions, and behaviors, as well as to study whether those differed across and within user demographics.

Researchers at MIT CSAIL and MIT LL conducted a pilot empirical survey to test drive the survey instrument in January 2021, and followed with a nationally representative survey in May 2021. The surveys focused on eliciting the engagement factors that public health professionals, as well as health care providers, might be able to influence. The survey methodology, data, and implications are detailed in program report ACTA-7 [22], and led to a CDC-MIT coauthored article in *Public Health Reports*. [35]

Study participants were generally willing to participate in contact tracing, quarantine, and notifying others through EN. They were less willing to download an EN app or notify others when the information being shared was visible to others. This confirmed EN implementers' initial assumptions that addressing user perception of communications, perceived involvement of different parties, and information visibility concerns should improve EN adoption and engagement. Most participants were unaware of currently available deployments, and requests by trusted health care or public health practitioners were the most likely to improve their likelihood of engagement with EN. The data also corroborated the hypothesis that users were more likely to use EN if it did not require them to install an app. Finally, the demographic and subdemographic analysis shed light on some seeming inconsistencies in prior research, where participant reported willingness differed between sub-demographics. The survey results were shared with CDC and other interested parties, in an effort to improve user adoption and PHA communication with citizens around matters of Exposure Notifications. The survey results can help organizations identify which subdemographic groups are receptive to which communication sources, as well as what to avoid.

While the EN implementation prioritized information security, adoption in many U.S. jurisdictions remains low. This research indicated that user perception may play a major role in user willingness to engage, where communications that are *perceived* as coming from a less trusted source may deter people from opting in. Future research into the perception of specific EN messaging is expected to yield additional insights. For instance, surveying a regional population could identify whether communication efforts actually reach and influence communities of interest. Conducting usability studies in a natural environment with different smartphone messages could yield insights into which messaging strategies are most effective. MIT's research indicates that removing workflow friction could improve user opt-in at a number of steps, such as providing enrollment quick response (QR) codes at entrances to indoor environments where people congregate to serve as a reminder (and easy enrollment opportunity) when and where the service is most needed. Finally, engaging with intended user communities to discuss perception of EN technologies and concerns during the pandemic, designing new interventions, and testing those during times of low transmission should help public health teams prepare for future infectious disease outbreaks.

8. ANALYSIS OF DEPLOYED EXPOSURE NOTIFICATION SYSTEMS

Although technologies like EN presented a significant opportunity for public health objectives during the COVID-19 pandemic, determining their population-level impact in regions where they were deployed is a complex question. MIT LL and MIT Sloan staff advised EN implementors and federal and state partners on the necessity of establishing metrics for EN performance without compromising the privacy of the individual EN users' data and identities. After the U.S. Exposure Notifications Private Analytics (ENPA) system was established in April 2021 [36] [37], MIT staff provided technical communications about ENPA and performed analysis of available data, in order to help others estimate how well EN deployments have performed in different jurisdictions, how much actual value EN has provided, and how the value of EN could be maximized.

At the close of Program 10383, there was still very little systematic data collected and shared by EN-implementing jurisdictions, despite the fact that, in principle, relevant data can be collected and published without impacting users' privacy. In particular, most available statistics on EN focused on the level of adoption in each jurisdiction. Other data that might illuminate how well EN actually detects new infectious cases, relative to other public-health interventions and to infection prevalence in the deployment environment, were not obtainable. Nevertheless, MIT devised a process flow model for EN, explored its predictions based on the available data from existing deployments, and showed how it could be used to predict the effects of varying initial conditions or policy changes within a jurisdiction. The process flow model can be extended to incorporate newly available ENPA features and other salient data that may be released in future systems like EN.

8.1 ENPA SYSTEM ARCHITECTURE

In order to provide public health authorities (PHAs) with metrics to help assess the effectiveness of the Exposure Notifications system, ENPA collects aggregate metrics while protecting privacy: individual user data is not revealed to any entity, and aggregate metrics are not revealed to Google or Apple. ENPA is enabled in ENX for PHAs that opt in to the analytics feature, and metrics include only those users who opt in to contribute data.

To achieve strong privacy guarantees, ENPA uses secure multi-party computation (MPC) and differential privacy. MPC is a type of cryptographic protocol that allows two or more parties to jointly compute a function of some data without actually seeing the data, instead operating on random "shares" of the data that reveal no information unless combined. Differential privacy requires that an aggregate metric not allow much to be inferred about any individual data contributor; this is achieved by adding random noise to the data and outputting approximate rather than exact statistics.

The ENPA system is executed by five parties: Google, Apple, the Internet Security Research Group (ISRG), the National Cancer Institute in the National Institutes of Health (NIH), and the MITRE

Corporation. At a high level, the system works as follows, as illustrated in Figure 19. First, Android and iOS phones add noise (for differential privacy) to their input data, generate cryptographic shares of the noised data, and encrypt those shares to the ISRG and NIH servers. The encrypted shares are sent to ingestion servers run by Google and Apple, which perform device attestation to ensure that only legitimate devices are contributing data before forwarding the encrypted shares to the ISRG and NIH servers. The ISRG and NIH servers then decrypt their shares and perform an MPC protocol to verify the validity of the inputs and obtain shares of the aggregate metrics, which are sent to the MITRE server. The MITRE server combines the final shares to obtain the aggregate metrics and makes them available through a web interface to the PHA.

Metrics provided by ENPA include, for each time period, estimates of how many users received exposure notifications, how many users interacted with (e.g., tapped or dismissed) these exposure notifications, how many users entered a verification code within 14 days after receiving an exposure notification, how many users shared their temporary exposure keys (TEKs) within 14 days after receiving an exposure notification, histograms of the number of days from an exposure to receiving a notification, and histograms of the computed risk scores (not only those resulting in exposure notifications).

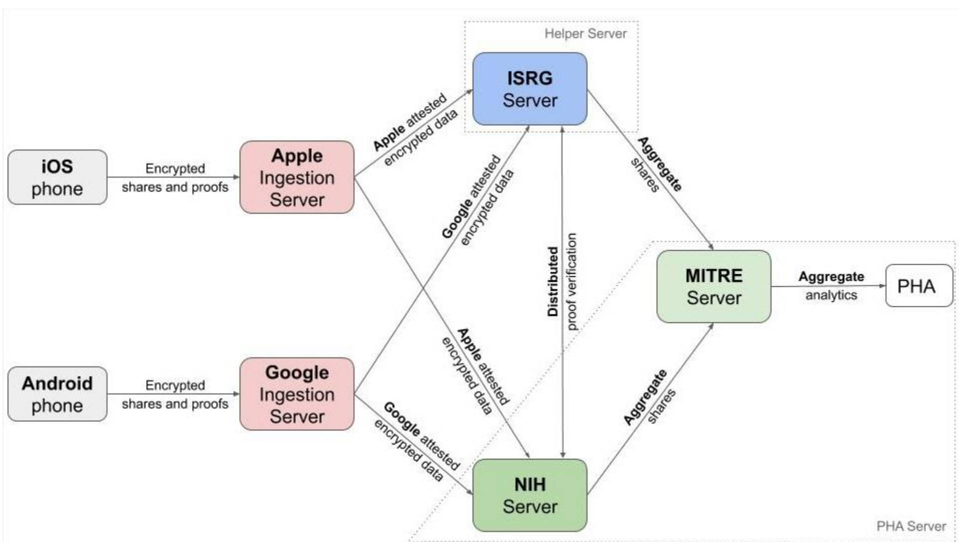


Figure 19. Exposure Notifications Private Analytics (ENPA) architecture. Source: [36]

8.2 EN PERFORMANCE ANALYSIS

8.2.1 Process Flow Model for EN

Most first-generation EN deployments lacked sufficient feedback loops, which are required for managing key trade-offs and maximizing the value and impact. However, in many jurisdictions, it was possible to obtain estimates of the number of concurrent EN users, the number of EN alerts issued per day, and/or the number of new cases among EN users, which became the starting point for the analysis. In order to work with key parameters from multiple jurisdictions in a consistent manner, MIT constructed a process model that represented a generic EN flow, as depicted in Figure 20.

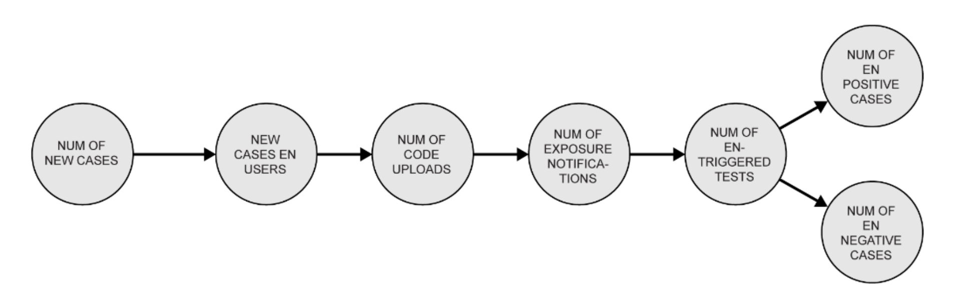


Figure 20. Generic EN process flow model.

Where the number of new cases among EN users was not recorded within the EN ecosystem, one could derive an estimate based on the EN adoption rate and the total number of new positive cases in a particular jurisdiction. From the number of new cases among EN users, and the average key upload rate, one could predict an expected number of EN validation codes issued and key uploads.[13] Each upload resulted in some number of triggered EN exposure notifications to other app users, determined by the risk score setting of the EN app at that jurisdiction and the behavior of the individuals involved. For simplicity's sake, the model assumed an average number of EN alerts per code/upload across the regional population.

[13] EN's implementation of "notify others" is a two-step process in which the user enters a personal verification *code*, followed by the *upload* of keys to a national server. Users may complete the first step but fail to complete the second. PHAs have access to metrics for the number of codes they have distributed to citizens, the number of codes claimed by EN users, and the number of successful key uploads. For simplicity, the process flow model combined the code claims and key uploads into a single "code uploads" estimate for the "notify others" step.

The number of EN users who received alerts determined the expected number of EN-triggered tests, based on the percentage of users in that jurisdiction who were likely to go and get tested following the receipt of an alert. Similarly, the model attempted to predict the expected number of help desk sessions triggered by EN alerts.

Figure 21 illustrates an assessment of the process flow values for the NHS COVID-19 app deployment in the UK, using metrics published by the NHS/Oxford/Turing team. [38] During the study period, 28% of the UK population was using the NHS app and 72% of app users who tested positive actually uploaded their EN keys. Each upload resulted, on average, in 4.4 new EN alerts to other app users. The study team estimated that only 50% of the app users who received an EN alert actually proceeded to get tested. Further, the study estimated that at least 6.1% of those tested after receiving an EN alert tested positive. Corresponding values from Switzerland (CH) and the Netherlands (NL) were assessed within the process flow model (included within Figure 23). [39] [40] [41] [42]

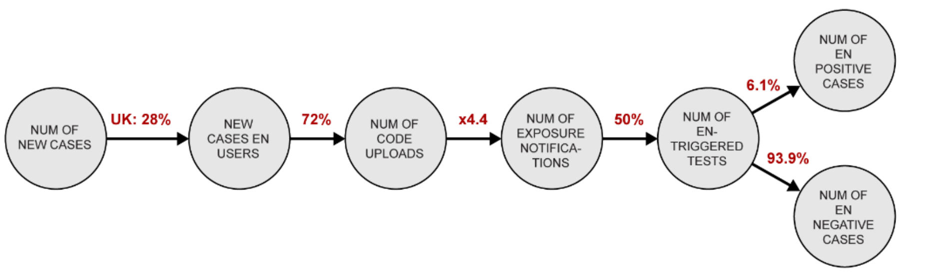


Figure 21. EN process flow model with UK deployment estimates.

This type of process flow abstraction and high-level estimations served as a basis for initial rough estimations and what-if analysis for various states and jurisdictions, which can then be further analyzed and refined. For example, applying the above UK-derived parameters to the number of estimated new cases in a particular U.S. jurisdiction on a given day, one could derive rough estimates as depicted in Figure 22. That is, given an estimated number of 1245 new cases in the state on a given day, based on the UK parameters, the state might require approximately 550 additional tests as a result of new EN alerts.

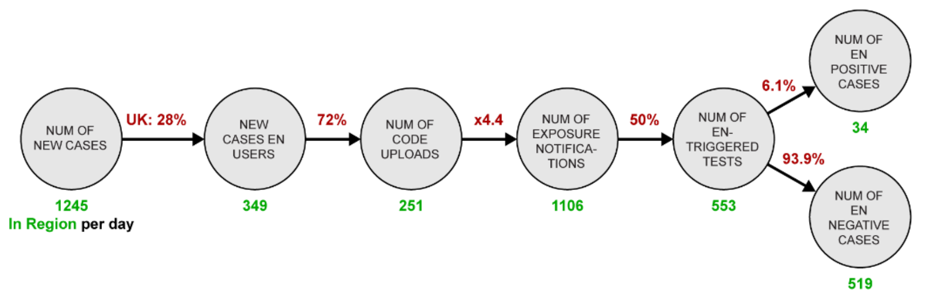


Figure 22. EN process flow model with projected estimates for a U.S. jurisdiction in Feb. 2021.

The most important parameter in the generic EN process flow was inspired by the epidemiological morbidity measure *secondary attack rate* (SAR), that is, the number of new cases among contacts of the index case. The corresponding concept within this analysis is the post-EN-positivity rate, or SAR_{EN}, which approximated the likelihood of a person receiving an EN alert to be infected. In the case of the UK study, the SAR_{EN} was estimated to be 6.1% (as discussed below, this actually represented a lower bound based on the way this estimation is derived).

8.2.2 Quantifying Overall Value of EN

In general, the higher the SAR_{EN}, the higher the public health value of the EN deployment. The SAR_{EN} within a given population, at a given time period, can vary from a minimum—equivalent to the general COVID-19 positivity rate within the population at that time, in which case the added value of EN is nil—to a maximum—corresponding to the true secondary attack rate of the dominant COVID-19 variant.

In order to assess the SAR_{EN} and the corresponding quantifiable value of each EN notification, the MIT model relied on the EN alert context, with attributes describing the alert-generation and alert-receipt events. To quantify the incremental benefits of EN and balance them with the incremental costs in a return-on-investment calculation, it was useful to classify post-EN-positive cases into four categories, as depicted in Figure 23:

1. EN alert recipient does not have symptoms and has not received a call from manual contact tracer (“EN-only asymptomatic”)

2. EN alert recipient does have symptoms and has not received a call from manual contact tracer (“EN-only symptomatic”)

3. EN alert recipient does not have symptoms and has received a call from manual contact tracer (“EN & MCT asymptomatic”)

4. EN alert recipient does have symptoms and has received a call from manual contact tracer ("EN & MCT symptomatic")

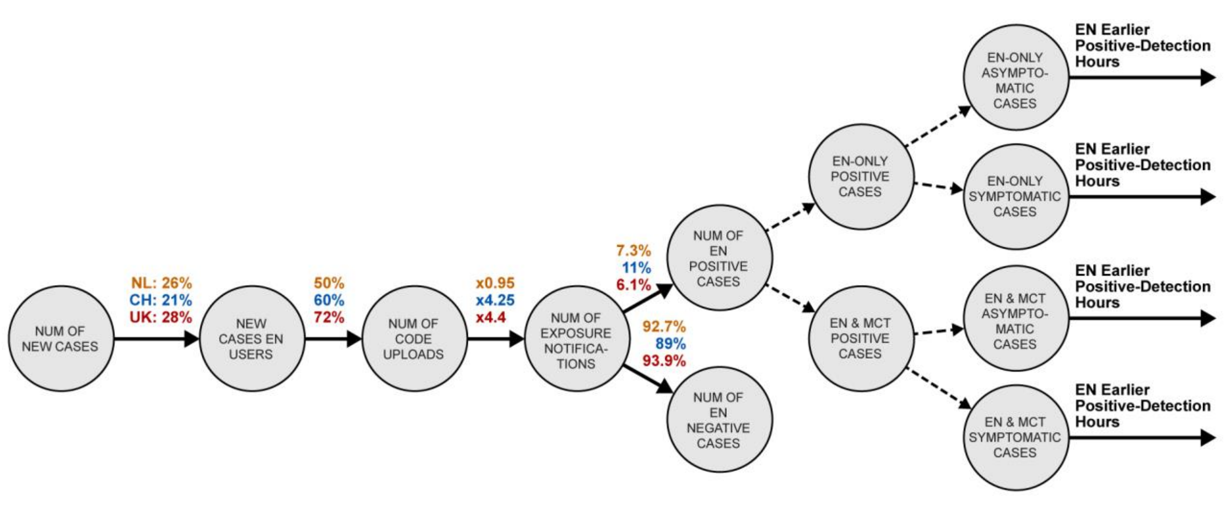


Figure 23. Classification of post-EN-positive cases, with upstream estimates from three European EN deployments [38].

Each category was expected to have a different level of incremental value. The incremental value in each case was conveyed by the earlier detection hours of the infection due to the EN alert, which depended on the timeline relationship between EN-triggered tests and tests triggered by other circumstances.

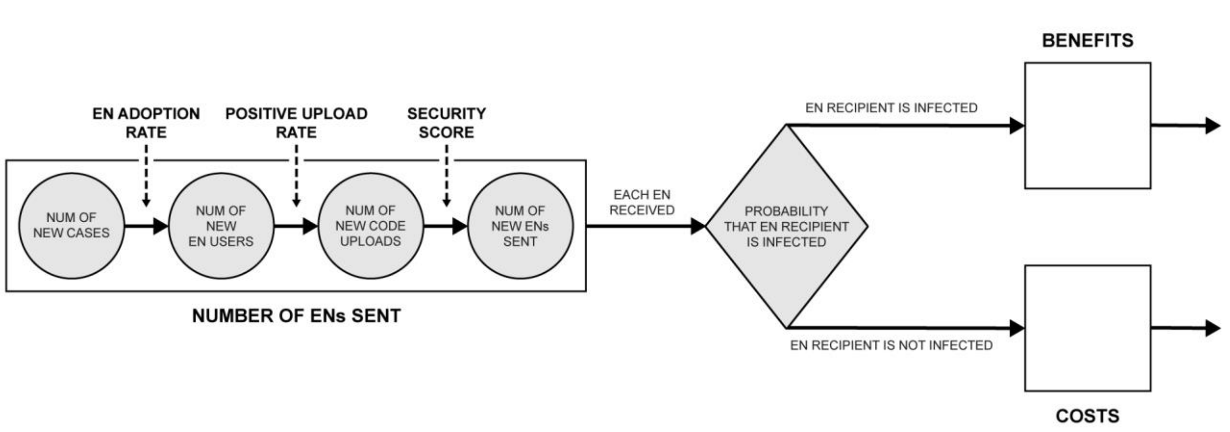


Figure 24. EN process flow components affecting cost-benefit ratio.

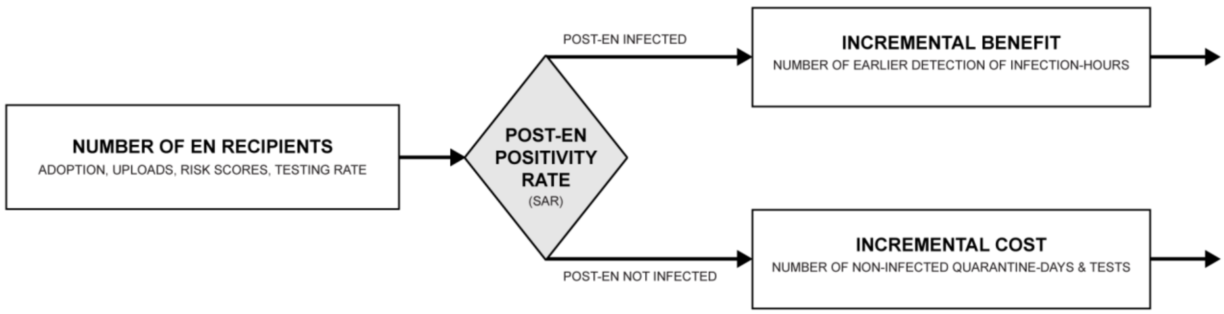


Figure 25. Incremental benefit and incremental cost.

The cost of EN in a given jurisdiction included the development, deployment, and operational costs of the EN solution in that jurisdiction. However, the total EN societal cost also included the costs associated with EN alerts to recipients who were *not* infected (Figure 24). In particular, as shown in Figure 25, such costs for EN "false positive" alerts included extra tests and extra quarantine days, with concomitant lost wages and productivity, reduced social interactions, and increased anxiety. Thus, the analysis defined the *incremental value* of each EN alert as:

the number of earlier detection hours and the public health value of each earlier detection hour (considering the four categories of cases, above) multiplied by SAR_{EN}, minus the cost of each false quarantine day and test overhead multiplied by ($1 - SAR_{EN}$).

The incremental value of each single EN notification, when multiplied by the number of notifications generated, produced an overall estimate of the value of the EN deployment. Predicting estimates of the deployment value allowed PHAs and implementers to explore, in a more systematic way, the tradeoffs between possible decisions they might make:

- Whether to deploy EN at all
- Which EN settings might satisfy the region's cost-benefit constraints
- Which policy changes might have positive or adverse effects on the overall value of the deployment.

An example of this type of analysis is shown in Figure 26, illustrating the cost-benefit relationship between three risk score settings: R1, R2, and R3. Each risk score was associated with a different number of notifications, and each notification derived a different overall value (cost-benefit balance). Under this scheme, R2 derived the highest benefit in the space where benefit exceeded cost (above and to the left of the dashed line).

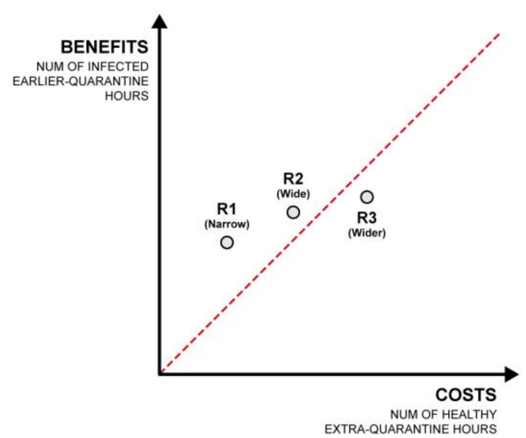


Figure 26. Predicted EN cost-benefit with notional risk score configurations R1, R2, and R3, for a given number of EN-user index cases.

To explore the effect of the risk score selection with a higher number of EN-user index cases (e.g., by increasing level of EN adoption), the analysis showed a proportional increase in the overall value (Figure 27). If a jurisdiction is debating whether to invest resources in methods that would increase adoption (e.g., outreach to targeted communities, partnerships with universities), the analysis could reassure them that the greater benefit is not outweighed by greater cost. In fact, the increase in cost would be lower than that associated with no adoption increase and a more sensitive, less specific EN configuration (R2).

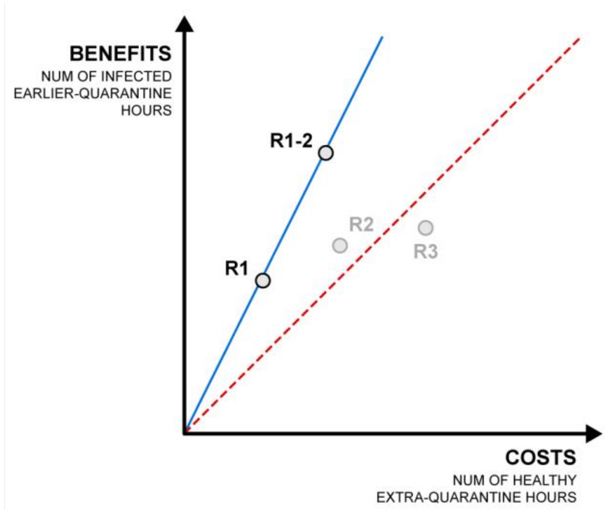


Figure 27. Predicted effect of doubling the number of EN-user index cases for R1, on average EN cost-benefit (R1-2).

The EN cost-benefits systematic analysis could also be the basis for evaluating proposed EN-specific or EN-adjacent policies. For example, in Figure 28, R’1 represents the value of EN using a policy in which recipients of an EN notification are required only to test, and then isolate only if the test is positive. The relaxation of the quarantine policy would significantly reduce the costs associated with quarantining uninfected EN alert recipients, while also somewhat reducing the benefits due to the delayed isolation start for infected individuals.

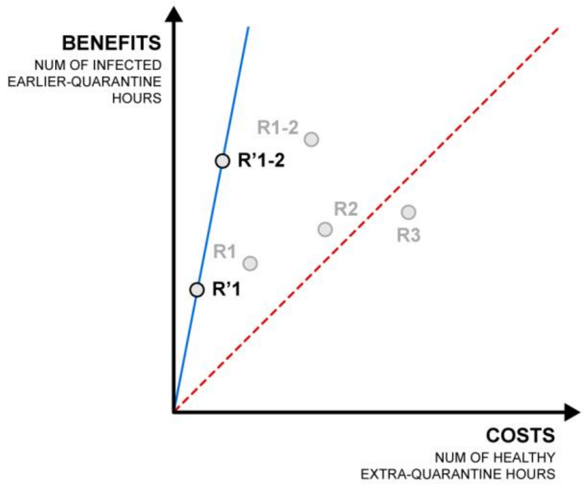


Figure 28. Predicted effect of test-to-isolate policy change for R1 and R1-2, on average EN cost-benefit (R'1 and R'1-2).

Similarly, policies regarding verified and non-verified results-based notification could be assessed within the same cost benefits reasoning structure. Such scenarios became increasingly significant with the prevalence of home testing and self-reporting. Figure 29 depicts such scenarios in which there are significantly enhanced benefits (speed, ease of use, etc.) and some potential increased risk (somewhat lower reliability of test results may produce a somewhat lower SAR_{EN}).

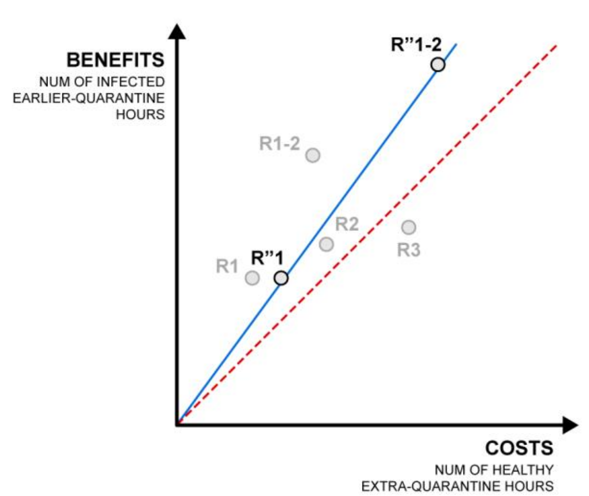


Figure 29. Predicted effect of relaxed verification policy change for R1 and R1-2, on average EN cost-benefit (R''1 and R''1-2).

8.2.3 Limitations

In general, despite the fact that the SAR_{EN} is the single most important fundamental parameter for assessing the performance, efficacy, and benefits of an EN solution, this value has not been available or derivable in most jurisdictions, and in the few jurisdictions for which it was assessed, only rough lower-bound approximations were obtainable.[14] In particular, in many jurisdictions, the only way to compute SAR_{EN} is with ENPA or ENPA-like solution. In the U.S., the ENPA data and the EN COVID Verification (ENCV) server data on codes claimed and keys shared are owned by each PHA and were not sufficiently available for analysis during the period of performance of Program 10383. In addition, states reported that when case rates were low, the noise added to ENPA aggregate metrics made analysis more challenging.

8.2.4 Suggested Future Work

MIT's metrics analysis approach has the potential to help PHAs understand SAR_{EN} for specific subpopulations of interest—e.g., those who are vaccinated against the virus. If ENPA were to be extended to allow EN users to opt into sharing their vaccination status/date/manufacturer, privately and anonymously, then more granular SAR_{EN} could be assessed. These could enable comparison of SAR_{EN} for EN notification recipients who are not vaccinated vs. those who are vaccinated, with respect to time since last dose and vaccine subtype (manufacturer or variant-specific boosters). This could help provide new insights on infectiousness within a particular region.

Another possible enhancement would allow for a more direct and accurate derivation of SAR_{EN} by including additional context data obtained in conjunction with the testing process (Figure 30).

[14] In such cases, the positivity rate is assessed by calculating the number of new positive cases which received an EN alert within the last 14 days, and dividing by the total number of alerts. However, this only produces a lower bound estimation because it does not consider the EN alert recipients who did not get tested after a notification, or did not upload their keys following positive test results.

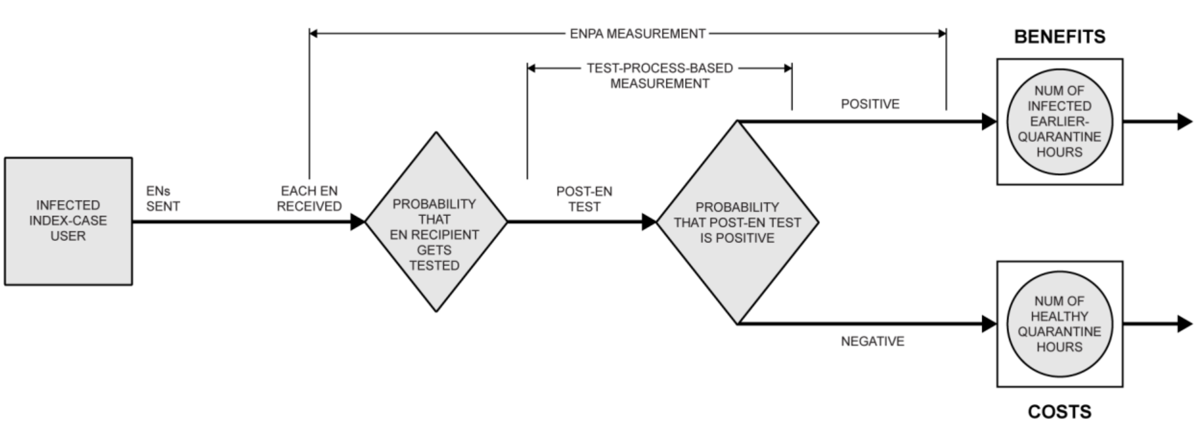


Figure 30. Proposed augmentation of EN process flow with test-time data collection, to derive post-notification positivity rates.

A few simple test-time questions, outlined in Table 5, could establish a richer EN alert context, and thereby enable the PHA to derive a more accurate SAR_{EN} and perform a more granular cost-benefits assessment. Such information collection and analysis could be performed in a manner that does not violate any of the privacy-by-design attributes of EN or ENPA. The questions could be incorporated into laboratory test metadata and/or into ENPA's metadata for self-reported positive tests.

Table 5

Suggested Test-Time Questions to Augment ENPA Metrics

In the last 14 days, did you:			
1. Receive an alert from EN?	Y/N	If Y, how many days ago?	How many days since exposure? (if known)
2. Receive any other notice of exposure?	Y/N	If Y, how many days ago?	How many days since exposure? (if known)
3. Experience any COVID-19 symptoms?	Y/N	If Y, how many days ago did symptoms start?	

Finally, tools should be developed to help PHAs to integrate the data available from ENCV, ENPA, and other sources, with the aim of calculating SAR_{EN} and examining trends over time. U.S. public health agencies have been reluctant to publish or even privately share their available metrics due to concerns about the misinterpretation of a highly nuanced and incomplete picture of EN's performance. A research framework based on mutual trust and collaboration could produce a much clearer picture of EN's performance throughout the pandemic. At the end of Program 10383's period of performance, a limited form of collaboration had taken shape with some U.S. states sharing data with a single private researcher, and the development of an open-source tool for combining and viewing ENCV and ENPA data from a single state at a time, shared with the Learning Lab participants. [43] Efforts such as these have been a late, but important, step in the right direction. These tools could be extended with the analysis approach described above to help states evaluate the costs and benefits of their EN decisions.

9. SUMMARY

The work conducted under Program 10383 validated the potential of AEN as a tool for extending the capacity and capabilities of conventional contact tracing efforts, especially as more transmissible variants of COVID-19 emerged, and as contact tracing resources were scaled back later in the pandemic. MIT LL and the PACT team members worked closely with staff at CDC, regional public health departments, and helth care experts to test and evaluate the flexibility of Exposure Notifications to meet the needs of different communities. Each component of the program ultimately assisted the local, national, and international communities of practice with technical exploration and decision support.

The data collected and models developed under this program suggest several possible applications to the problems of estimating potential exposure to other airborne infectious pathogens. However, technical challenges remain. The risk estimate could be made more specific by incorporating info about mask wearing and the environment of the exposure, personal susceptibility factors, and variant prevalence. The RF sensor could be improved with more rigorous calibration and new hardware, without sacrificing privacy. Disease surveillance methods could be enhanced with higher-quality aggregated and anonymized data willingly shared by AEN users. Crucially, a better feedback loop should be devised such that metrics for the number of notifications and the secondary attack rate (SAR_{EN}) can be tracked, and can inform decisions regarding sensor-driven notification settings (e.g., risk scoring configurations) and post-notification policy decisions (e.g., whether to quarantine or test following a notification). The key feedback parameters should be tracked continuously to detect evolving circumstances as early as possible and to make timely adjustments.

The human factors in an AEN system are as important to its success as the technical factors. From this work, we understand more clearly what benefits people expect from using an AEN system, and what personal information they think is worth sharing in a public health crisis. Exposure Notification and other private automatic contact tracing systems demonstrated that individual citizens and smartphone technology can help with disease surveillance and mitigation. However, we identified multiple missed opportunities for better communication and justification of the value of AEN, both to individual citizens and to political and administrative decision makers, through data-driven cost-benefits analyses.

Exposure notification technologies offer a tremendous promise to transform important aspects of public health globally. Improved feedback, in a privacy preserving manner, can serve as a basis for managing important trade-offs, assessing the costs and benefits, and maximizing the overall value. Designed appropriately, and assessed appropriately, innovative automated exposure-notification technologies can offer tremendous benefits for the future of public-health.

This page intentionally left blank.

APPENDIX A: STATE X PILOT 0 TEST SHEET

Test Summary Data Sheet:

Test	Objective (Location)	Hardware/ OS	Criteria For Success	Pass?
0	Successful activation of ENX (home)	iPhone 13.7 iPhone 14.3 Android 8 Android 10 Android 11	1. Able to perform opt-in on all phones 2. Beacons observed from all phones on at least one Android 3. Able to see exposure check on all phones, with no alerts received 4. Able to erase beacon history on all phones	
1	Successful key upload from iOS 13.7 and exposure check on Android 8/10/11 (ASDF)	iPhone 13.7 Android 8 Android 10 Android 11	1. Able to issue verification code by text message and enter on iPhone with confirmation of key upload 2. Able to see exposure check on all phones, with alerts received on all 3 exposed phones	
2	Successful key upload from iOS 14.3 and exposure check on Android 8/10/11 (ASDF)	iPhone 14.3 Android 8 Android 10 Android 11	1. Able to issue verification code by text message and enter on iPhone with confirmation of key upload 2. Able to see exposure check on all phones, with alerts received on all 3 exposed phones	
3	Successful exposure check across both iPhone versions (home)	iPhone 13.7 iPhone 14.3	1. Able to enter verification code on both iPhones with confirmation of key upload 2. Able to see exposure check on both iPhones, with alerts received on both.	
4	Successful key upload from Android 8 and exposure check on iOS 13.7/14.3, Android 10/11 (ASDF)	iPhone 13.7 iPhone 14.3 Android 8 Android 10 Android 11	1. Able to issue verification code by text message and enter on Android with confirmation of key upload 2. Able to see exposure check on all phones, with alerts received on all 4 exposed phones	
5	Successful key upload from Android 11 and exposure check on iOS 13.7/14.3, Android 8/10 (ASDF)	iPhone 13.7 iPhone 14.3 Android 8 Android 10 Android 11	1. Able to issue verification code by text message and enter on Android with confirmation of key upload 2. Able to see exposure check on all phones, with alerts received on all 4 exposed phones	

If the setting for exposure check interval is 4 hours, we estimate the first 6 tests will take between 2 and 3 days.

Additional tests if time permits:

6	Successful key upload from Android 10 and exposure check on iOS 13.7/14.3, Android 8/11 (ASDF)	iPhone 13.7 iPhone 14.3 Android 8 Android 10 Android 11	1. Able to issue verification code by text message and enter on Android with confirmation of key upload 2. Able to see exposure check on all phones, with alerts received on all 4 exposed phones	
7 and up	Exposure checks for keys at ranges targeting different attenuation "buckets" (ASDF)	iPhone 14.3 Android 8 Android 10 Android 11	1. After exposure at ranges beyond O, M, and N thresholds for attenuation, and for short durations, able to see exposure check on all phones, *without* alerts received on all 4 exposed phones	

Notes on test locations and spacing

Test 0 may be performed at home, a representative environment for ENX users to activate ENX and perform exposure checks. This also will reduce the total number of days needed for testing. Phones will be located no more than a few feet apart, running ENX, for the number of hours it takes for all to have performed at least one exposure check (may be up to 5 hours).

Test 3 may be performed at home with family members, rather than Phantom body analogues, also in the interest of time. Phones will be held in the hand, seated 4 feet apart and facing each other, for 20 minutes. Phones will then be separated beyond Bluetooth range for key upload and exposure check.

ASDF tests (1, 2, 4, 5) will be performed with phones at 3' separation from the "sick" phone, for 30 minutes duration, to generate a "strong" exposure and a "true positive" test case. Test 0 is one form of "true negative" (exposure occurs, but nobody is sick, so no alerts should be observed). Tests 7 and up are intended to exercise the other form of "true negative" (exposure occurs, but not too close or too long). If we have time to perform tests beyond #6, we will use the debug-level attenuation information recorded on the Android phones during tests 0-6 to inform the physical spacing in tests 7 and up.

Engineering data to be collected
1. Android ENS service-level logs (RSSI, timestamps, encrypted and cleartext metadata)
2. Android screenshots of key upload times, exposure check times and results
3. iPhone screenshots of key upload times, exposure check times and results
4. Physical configuration (spacing and hardware) -- human notes

Deliverables
1. Completed test summary data sheet with GO / NO GO recommendation
2. Photos of test configurations

APPENDIX B: STATE X PILOT 0 RESULTS

Pilot 0 Snapshot

	Organizational Lead	# of phones	Phone users	Proximity Conditions	Approach	Functionality Demonstrated	Dates	Status
Pilot 0	MIT LL	9	None (robots, lab benches, etc.)	Benchtop & Test Range, Home Envt.	• Basic functionality testing • Debugging • Fake verifications • Fake alerts • No X-IDS connectivity • APHL MVS connectivity (manual) • APHL NKS connectivity	• Configuration parameters properly set • APHL MVS connectivity established and providing basic functionality • APHL NKS uploads working with short codes	Week 1	PASS
Pilot 0 follow-on	MIT LL & State X DPH	9	None (robots, lab benches, etc.)	Benchtop & Test Range, Home Envt.	• SMS functionality testing • CSV batch upload testing	• APHL MVS providing SMS functionality • APHL MVS batch SMS working • APHL NKS uploads working with deep links	Week 3	PASS

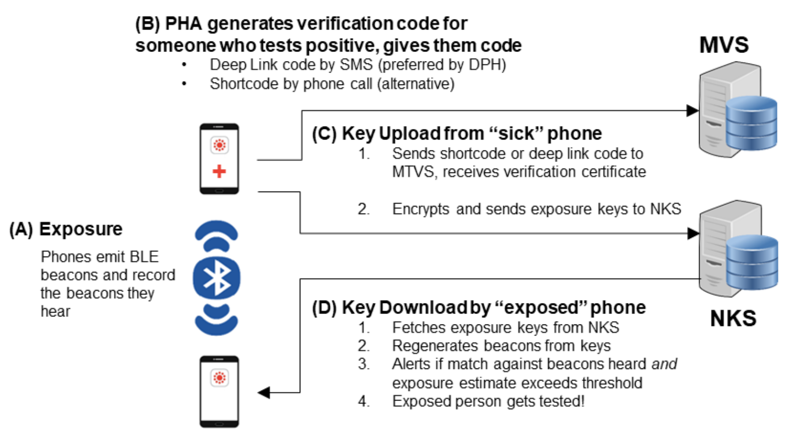
- **SMS capability enabled during gap week (Week 2)**
- **MIT LL verified X-Alert configuration is deployed as submitted to Apple|Google**
- **iOS 13.7 could not be tested during Pilot 0 or 1 due to lack of support for Profile and Ramp Modes in OS**
- **Screenshots, video captures, and error message workflows captured in support of Help Desk staff training**

Pilot 0 Results - 1

APHL: Assn. of Public Health Labs
CSV: a file format
"X-Alert": Exposure Notifications Express soln.
"X-IDS": Infectious disease surveillance database
MVS: Multi-tenant Verification Server
NKS: National Key Server

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Exposure Notification CONOPS



Glossary

Beacon: Broadcast message containing randomized text "identifier" and signal strength

BLE: Bluetooth Low Energy

Deep Link: HTTPS link which opens in X-Alert and begins the sharing workflow when clicked; includes the one-time use code for verification

MVS: Multi-tenant Verification Server, hosted by APHL

NKS: National Key Server, hosted by APHL

Shortcode: 8-digit one-time use code for verification

SMS: Short Message Service, a text messaging service

Pilot 0 Results - 2

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Pilot 0 Objectives

- Does the system work as expected?
  - Can the user join, pause, and leave X-Alert?
  - Can the user enter codes, upload keys, and receive alerts?
- Is information clearly conveyed? Do links work?
- Does the system *fail* as expected?
  - Can the user understand why an attempted action failed?
  - Does X-Alert recommend the correct action to fix the situation?
  - Does the user require help from DPH to proceed? (Should they?)

Example:

User attempts to enter a short code for key upload, but server is unreachable.

Ideally, phone should tell them that the network is the issue, and they should succeed with re-entering code after enabling or improving their network connection.

Test Matrix: Basic Functionality

	Opt In Works	Opt Out Works	Sends Beacons	Checks For Exposures	Reasonable Check Interval	No Alert When No Sick Phone	"Erase History" Works	Uninstall Works	Pause Works
iOS 14.3	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
iOS 14.4	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
Android 8.1	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS*	PASS
Android 10	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
Android 11	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

* Uninstall failed silently once on Android 8.1, but succeeded on second try

Test Matrix: Short Code Verification and Key Upload

	Short Code Key Upload Works	Short Codes Expire (15 min)	Short Codes Fail Visibly (EN off or no network)	Short Codes Work After EN Fail	Short Codes Work After Network Fail
iOS 14.3	PASS	PASS	PASS	MIXED[1]	NO[3,4]
iOS 14.4	PASS	PASS	PASS	MIXED[2]	NO[3,4]
Android 8.1	PASS	PASS	PASS	n/a Android directs user to enable EN before they can enter code	PASS
Android 10	PASS	PASS	PASS		PASS
Android 11	PASS	PASS	PASS		PASS

"EN Fail": API call to release keys fails because EN is off

"Network Fail": wifi and cellular data off; airplane mode; wifi/cellular on but poor connection

(1) Short code was claimed, but phone reported an error and told user to contact PHA. We learned later that this was an artifact of repeat tests with a single phone on one day. When test was repeated on another day, phone said upload succeeded without Bluetooth, and exposed phone alerted.

(2) Phone said upload succeeded; observing phone had key match, but surprisingly no alarm. When test was repeated, phone said upload succeeded without Bluetooth, and exposed phone alerted.

(3) When network was enabled but connection was poor, phone showed a progress wheel and eventually recommended to try again later. Code was unclaimed on server, but phone would not retry the same short code after network connectivity improved.

(4) When network was disabled at first attempt, then re-enabled, second attempt still told user to turn on wi-fi or cellular.

These are non-blocking NO cases: Pilot 1 may proceed with DPH informed of expected behavior

Pilot 0 Results - 5

EN: Exposure Notifications, as available through "X-Alert" interface

LINCOLN LABORATORY
Massachusetts Institute of Technology

Test Matrix: Deep Link Verification and Key Upload

	Deep Link Key Upload Works	Deep Links Expire (15 min)	Deep Links Fail Visibly (EN off or no network)	Deep Links Work After EN Fail	Deep Links Work After Network Fail
iOS 14.3	PASS	PASS	PASS	MIXED[2]	MIXED[3,4]
iOS 14.4	PASS	PASS	PASS	PASS	PASS
Android 8.1	PASS	PASS	PASS	n/a Android directs user to enable EN before they can enter code	PASS
Android 10	PASS	PASS	PASS		PASS
Android 11	PASS	PASS	MIXED[1]		PASS

"EN Fail": API call to release keys fails because EN is off

"Network Fail": wifi and cellular data off; airplane mode; wifi/cellular on but poor connection

(1) When EN off, phone asked to turn on EN, but the button it provided to do that had no effect (v11 only). Closed the view and enabled EN within app settings. Could not reproduce problem behavior later the same day.

(2) Have seen phone "successfully share" with Bluetooth both enabled and disabled; have also seen nonspecific error message with Bluetooth disabled, and link worked after enabling it.

(3) No wifi / no cellular test: Had to close tab in Safari after deeplink failed to reach server, to get past Safari error; retry failed and needed new link.

(4) "Airplane+wifi" mode worked for deeplink

These are non-blocking NO cases: Pilot 1 may proceed with DPH informed of expected behavior

Pilot 0 Results - 6

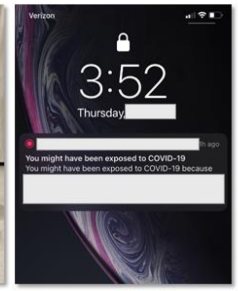
EN: Exposure Notifications, as available through "X-Alert" interface

LINCOLN LABORATORY
Massachusetts Institute of Technology

Test Matrix: EN Alerts

		Alerted after 30 minute exposure at 2 feet separation				
		iOS 14.3	iOS 14.4	Android 8.1	Android 10	Android 11
"Sick" phone	iOS 14.3	n/a	PASS	PASS	PASS	PASS
	iOS 14.4	PASS	n/a	PASS	PASS	PASS
	Android 8.1	PASS	PASS	n/a	PASS	PASS
	Android 10	PASS	PASS	PASS	n/a	PASS
	Android 11	PASS	PASS	PASS	PASS	n/a

- Did not test same-OS exposures (we only have one each of Android 8.1 & 11, and cross-version test cases are more likely to have issues than same-version)



Pilot 0 Results - 7

EN: Exposure Notifications, as available through "X-Alert" interface

LINCOLN LABORATORY
Massachusetts Institute of Technology

GLOSSARY

ACTA	Automated Contact Tracing Assessments
AEN	Automated Exposure Notification
APHL	Association of Public Health Laboratories
API	Application programming interface
ASDF	Autonomous Systems Development Facility
Attenuation	Reduction of signal amplitude
BLE	Bluetooth Low Energy
BLEMUR	Bluetooth Low Energy Model of User Risk
CDC	Centers for Disease Control and Prevention (U.S.)
COVID-19	Coronavirus disease caused by the SARS-CoV-2 virus
DPH	Department of Public Health
EN	Exposure Notifications
ENCV	Exposure Notification COVID Verification
ENPA	Exposure Notifications Private Analytics
ENX	Exposure Notifications Express
EPFL	École polytechnique fédérale de Lausanne
FDR	False detection rate
GAEN	Google-Apple Exposure Notifications
HCI	Human-computer interaction
HF	Human factors
ISRG	Intenet Security Research Group
LFPH	Linux Foundation Public Health

MCT	Manual contact tracing (i.e., not automated)
MIT	Massachusetts Institute of Technology
MIT CSAIL	MIT Computer Science and Artificial Intelligence Laboratory
MIT IPRI	MIT Internet Policy and Research Initiative
MIT LL	MIT Lincoln Laboratory
MITRE	The MITRE Corporation
MPC	Multi-party computation
MVS	Multi-tenant Verification Server (U.S.)
NHS	National Health Service (U.K.)
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NKS	National Key Server (U.S.)
OS	Operating system
P(Alert)	Probability of EN alert occurring
P(D)	Probability of detection
P(FA)	Probability of false alarm
PACT	Private Automated Contact Tracing
PHA	Public health authority
RF	Radio frequency
RPI	Rolling Proximity Identifier, a short-lived token generated from the TEK
RSSI	Received Signal Strength Indicator *or* Risk Scoring Symposium Inivitational
SAR	Secondary attack rate, the number of new cases among contacts of the index case
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2

SimAEN	Simulation of Automated Exposure Notification
SMS	Short Message Service
TEK	Temporary Exposure Key, a cryptographic token generated on the smartphone once per day
UTC	Universal Time Coordinates
USB	Universal Serial Bus
UX	User experience
WHO	World Health Organization
X-Alert	State X’s ENX implementation (anonymized)
X-IDS	State X’s infectious disease surveillance database (anonymized)

This page intentionally left blank.

REFERENCES

[1] R. L. Rivest, H. Abelson, J. Callas, R. Canetti, K. Esvelt, D. K. Gillmor, et al., "The PACT Protocol Technical Specification," 08 04 2020. [Online]. Available: https://pact.mit.edu/wp-content/uploads/2020/11/The-PACT-protocol-specification-2020.pdf. [Accessed 01 06 2022].

[2] Google, "Exposure Notification API launches to support public health agencies," 20 05 2020. [Online]. Available: https://blog.google/inside-google/company-announcements/apple-google-exposure-notification-api-launches/. [Accessed 01 05 2020].

[3] United States Government Accountability Office, "Exposure Notification: Benefits and Challenges of Smartphone Applications to Augment Contact Tracing," GAO-21-104622, Sept. 2021.

[4] 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, June 2022. doi: 10.1177/00333549221099533.

[5] Google, "Exposure Notifications BLE attenuations," [Online]. Available: https://developers.google.com/android/exposure-notifications/ble-attenuation-overview. [Accessed 10 April 2022].

[6] Centers for Disease Control and Prevention, "CDC Museum COVID-19 Timeline," [Online]. Available: https://www.cdc.gov/museum/timeline/covid19.html. [Accessed 19 04 2022].

[7] World Health Organization, "Tracking SARS-CoV-2 variants," 12 04 2022. [Online]. Available: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. [Accessed 19 04 2022].

[8] C. Zimmer, E. Anthes and A. Jacobs, "Omicron: What We Know About the Dominant Coronavirus Variant," The New York Times, 01 04 2022.

[9] CDC, "Contact Tracing : Part of a Multipronged Approach to Fight the COVID-19 Pandemic," [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/php/principles-contact-tracing.html. [Accessed 30 April 2020].

[10] Apple and Google, "Exposure Notification Bluetooth Specification," 20 April 2020. [Online]. Available: https://www.blog.google/documents/62/Exposure_Notification_-_Bluetooth_Specification_v1.1.pdf/. [Accessed 30 April 2022].

[11] Apple and Google, "Exposure Notification Cryptography Specification," April 2020. [Online]. Available: https://blog.google/documents/69/Exposure_Notification_-_Cryptography_Specification_v1.2.1.pdf/. [Accessed 30 April 2022].

[12] Association of Public Health Laboratories, "Exposure Notifications," [Online]. Available: https://www.aphl.org/programs/preparedness/Crisis-Management/COVID-19-Response/Pages/exposure-notifications.aspx. [Accessed 22 June 2022].

[13] Apple, "ENExposureConfiguration | Apple Developer Documentation," [Online]. Available: https://developer.apple.com/documentation/exposurenotification/enexposureconfiguration. [Accessed 22 June 2022].

[14] M. C. Schiefelbein, R. C. Gervin Jr., J. St. Germain and S. Mazzola, "Bluetooth Low Energy (BLE) Data Collection for COVID-19 Exposure Notfication," MIT Lincoln Laboratory, ACTA-2, Apr. 2022.

[15] M. C. Schiefelbein, R. C. Gervin Jr., J. St. Germain and S. Mazzola, "COVID-19 Exposure Notification in Simulated Real-World Environments," MIT Lincoln Laboratory, ACTA-3, Apr. 2022.

[16] G. Gettliffe, "Modeling Probability of Alert of Bluetooth Low Energy-Based Automatic Exposure Notifications," MIT Lincoln Laboratory, ACTA-4, Apr. 2022.

[17] 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.

[18] J. Wanger, "COVID-19 Exposure Notification Express vs. Custom App: Which to Choose?," 29 Sept. 2020. [Online]. Available: https://www.lfph.io/2020/09/29/enx-vs-custom/. [Accessed 22 June 2022].

[19] Chatham House, "Chatham House Rule | Chatham House – International Affairs Think Tank," 2022. [Online]. Available: https://www.chathamhouse.org/about-us/chatham-house-rule. [Accessed 30 04 2022].

[20] Commonwealth of Pennsylvania, "Gov. Wolf Encourages Pennsylvanians to Download and Share COVID Alert PA App, More than 70,000 Downloads Since Launch," 24 09 2020. [Online]. Available: https://www.governor.pa.gov/newsroom/gov-wolf-encourages-pennsylvanians-to-download-and-share-covid-alert-pa-app-more-than-70000-downloads-since-launch/. [Accessed 24 06 2022].

[21] E. H. Shen, B. Pelletier, J. E. Riordan, R. L. Rivest, W. W. Streilein, J. D. Wilkinson and M. A. Zissman, "Exposure Notification Privacy and Security Analysis," MIT Lincoln Laboratory, ACTA-1, May 2022.

[22] J. D. Alekseyev, I. Liccardi, V. L. A. Woltz and M. E. Zurko, "Toward EN Adoption: Bridging the Gap Between Stated Intention and Actual Use," MIT Lincoln Laboratory, ACTA-7, June 2022.

[23] J. Benzler, M. Briers, M. Flowers, S. Halai, B. Harris, B. T. Karras, et al., "Configuring Exposure Notification Risk Scores v1.0.0" [Online]. Available: https://github.com/lfph/gaen-risk-scoring/blob/d06da88fe1912529f0e6c6e6c0c3ca91bd5d93fc/risk-scoring.md. [Accessed 10 04 2022].

[24] E. Aronoff-Spencer, J. Benzler, T. Burner, C. Fraser, A. Higgins, B. T. Karras, et al., "Configuring Exposure Notification Risk Scores for COVID-19 v2.0," 11 12 2021. [Online]. Available: https://github.com/lfph/gaen-risk-scoring/blob/2.0.0/risk-scoring.md. [Accessed 10 04 2022].

[25] MIT Lincoln Laboratory, "Autonomous Systems Development Facility," [Online]. Available: https://www.ll.mit.edu/about/facilities/autonomous-systems-development-facility. [Accessed 10 04 2022].

[26] Google, "GitHub - google/exposure-notifications-internals," [Online]. Available: https://github.com/google/exposure-notifications-internals . [Accessed 10 04 2022].

[27] MIT Lincoln Laboratory, "GitHub - mitll/PACT-Exposure-Notification-Beacons," 27 09 2021. [Online]. Available: https://github.com/mitll/PACT-Exposure-Notification-Beacons.

[28] MIT Lincoln Laboratory, "GitHub - mitll/Exposure-Visualization-Tool," 05 01 2022. [Online]. Available: https://github.com/mitll/Exposure-Visualization-Tool.

[29] Google, "Exposure Notifications BLE Calibration Calculation," [Online]. Available: https://developers.google.com/android/exposure-notifications/ble-attenuation-computation. [Accessed 10 04 2022].

[30] MIT Lincoln Laboratory, "Lincoln Laboratory Supercomputing Center," [Online]. Available: https://www.ll.mit.edu/about/facilities/lincoln-laboratory-supercomputing-center. [Accessed 29 06 2022].

[31] J. Alekseyev, "Development and Validation of the Public-Facing SimAEN Web Application," MIT Lincoln Laboratory, ACTA-6, June 2022.

[32] W. Streilein, L. Finklea, D. Schuldt, M. C. Schiefelbein, R. Yahalom and A. Norige, "Evaluating COVID-19 Exposure Notification Effectiveness With SimAEN: A Simulation Tool Designed for Public Health Decision Making," Public Health Reports, August 2022. doi: 10.1177/00333549221116361.

[33] MIT Lincoln Laboratory, "SimAEN," July 2021. [Online]. Available: https://simaen.philab.cdc.gov/. [Accessed 29 June 2022].

[34] 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 29 June 2022].

[35] I. Liccardi, J. Alekseyev, V. L. Woltz, J. E. McLean and M. E. Zurko, "Public willingness to engage with COVID-19 contact tracing, quarantine, and exposure notification," Public Health Reports, Oct. 2022, doi: 10.1177/00333549221125891.

[36] Apple and Google, "Exposure Notification Privacy-preserving Analytics (ENPA) White Paper," April 2021. [Online]. Available: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf. [Accessed 02 06 2022].

[37] Google, "Analytics in Exposure Notifications Express: FAQ," [Online]. Available: https://github.com/google/exposure-notifications-android/blob/master/doc/enexpress-analytics-faq.md. [Accessed 02 06 2022].

[38] C. Wymant, L. Ferretti, D. Tsallis, et al., "The epidemiological impact of the NHS COVID-19 app," Nature, vol. 594, pp. 408-412, 2021.

[39] D. Menges, H. Aschmann, A. Moser, C. L. Althaus and V. von Wyl, "The role of the SwissCovid digital proximity tracing app during the pandemic response: results for the Canton of Zurich," 3 2 2021. [Online]. Available: https://www.medrxiv.org/content/10.1101/2021.02.01.21250972v1. [Accessed prior to Sept 1, 2021].

[40] P. Daniore, T. Ballouz, D. Menges and V. von Wyl, "The SwissCovid Digital Proximity Tracing App after one year: Were expectations fulfilled?," Swiss Med Wkly, vol. 151, no. w30031, 2021.

[41] W. Ebbers, L. Hooft, N. van der Laan and E. Metting, "Evaluatie CoronaMelder: een overzicht na 9 maanden," Tilburg University, 2021.

[42] Government of the Netherlands, "CoronaMelder data dashboard," [Online]. Available: https://coronamelder.nl/en/faq/3-2-coronamelder-data-dashboard/. [Accessed Feb 2021].

[43] B. Pugh, "GitHub - billpugh/GAEN-Analytics: A tool to allow public health authorities that have deployed the ENX version of GAEN to access and analyze their ENPA and ENCV metrics.," 31 05 2022. [Online]. Available: https://github.com/billpugh/GAEN-Analytics. [Accessed 28 06 2022].

[44] Bundesamt für Statistik (BFS) - Experimentelle Statistiken, "Swiss Covid Proximity Tracing App Monitoring | FSO - Experimental statistics," [Online]. Available: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.html.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
17 November 2022

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Automated Contact Tracing Assessment

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

M.C. Schiefelbein, C. Ishikawa, E.H. Shen, R. Yahalom, and A.S. Norige

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 02421-6426

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-1287

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Centers for Disease Control and Prevention

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.

13. SUPPLEMENTARY NOTES

13. ABSTRACT
In July 2020, CDC's Innovative Technologies Team designated MIT LL and the PACT team as trusted technical advisors on the deployment of private automated contact tracing systems as part of its overall public health response to COVID-19. The Innovative Technologies Team sought to answer the following key question regarding automated contact tracing: Does automated contact tracing have sufficient public health value that it is worthwhile to integrate it at scale into existing and evolving manual contact tracing systems? MIT LL focused the assessment efforts on the systems being built and deployed. There were two immediate and significant challenges to observing and quantifying the performance of the system as a whole: first, the privacy preserving design decisions of PACT and the system implementers denied access to system-level performance metrics, and second, obtaining accurate "ground truth" data about risky encounters in the population, against which to measure the detector performance, would require an unacceptable level of effort and intrusion. Therefore, MIT LL designed a set of parallel research activities to decompose the problem into components that could be assessed quantifiably (Bluetooth sensor performance, algorithm performance, user preferences and behaviors), components that could be assessed qualitatively (potential cybersecurity risks, potential for malicious use), and components that could be modeled based on current and emergent knowledge (population-level effects).

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**			**19b. TELEPHONE NUMBER** *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	None	88	

This page intentionally left blank.