frontiers in Digital Health

POLICY AND PRACTICE REVIEWS
published: 05 August 2021
doi: 10.3389/fdgth.2021.677929

Toward a Common Performance and Effectiveness Terminology for Digital Proximity Tracing Applications

Wouter Lueks[1], Justus Benzler[2], Dan Bogdanov[3], Göran Kirchner[2], Raquel Lucas[4], Rui Oliveira[5], Bart Preneel[6], Marcel Salathé[7], Carmela Troncoso[1†] and Viktor von Wyl[8,9*†]*

[1] Security and Privacy Engineering Laboratory, School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [2] Robert Koch Institute, Berlin, Germany, [3] Cybernetica AS, Tartu, Estonia, [4] Medical School and Institute of Public Health (EPIUnit), Universidade Do Porto, Porto, Portugal, [5] Institute for Systems and Computer Engineering, Technology and Science & University of Minho, Porto, Portugal, [6] Department of Electrical Engineering, Katholieke Universiteit Leuven and IMEC, Leuven, Belgium, [7] Digital Epidemiology Laboratory, School of Life Sciences, School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, Global Health Institute, Geneva, Switzerland, [8] Digital and Mobile Health Group, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland, [9] Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Yun William Yu, University of Toronto Scarborough, Canada
Milena B. Cukic, Amsterdam Health and Technology Institute (AHTI), Netherlands

***Correspondence:**
Viktor von Wyl
viktor.vonwyl@uzh.ch
Carmela Troncoso
carmela.troncoso@epfl.ch

† These authors share last authorship

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *08 March 2021*
Accepted: *08 July 2021*
Published: *05 August 2021*

Citation:
Lueks W, Benzler J, Bogdanov D, Kirchner G, Lucas R, Oliveira R, Preneel B, Salathé M, Troncoso C and von Wyl V (2021) Toward a Common Performance and Effectiveness Terminology for Digital Proximity Tracing Applications. Front. Digit. Health 3:677929. doi: 10.3389/fdgth.2021.677929

Digital proximity tracing (DPT) for Sars-CoV-2 pandemic mitigation is a complex intervention with the primary goal to notify app users about possible risk exposures to infected persons. DPT not only relies on the technical functioning of the proximity tracing application and its backend server, but also on seamless integration of health system processes such as laboratory testing, communication of results (and their validation), generation of notification codes, manual contact tracing, and management of app-notified users. Policymakers and DPT operators need to know whether their system works as expected in terms of speed or yield (performance) and whether DPT is making an effective contribution to pandemic mitigation (also in comparison to and beyond established mitigation measures, particularly manual contact tracing). Thereby, performance and effectiveness are not to be confused. Not only are there conceptual differences but also diverse data requirements. For example, comparative effectiveness measures may require information generated outside the DPT system, e.g., from manual contact tracing. This article describes differences between performance and effectiveness measures and attempts to develop a terminology and classification system for DPT evaluation. We discuss key aspects for critical assessments of whether the integration of additional data measurements into DPT apps may facilitate understanding of performance and effectiveness of planned and deployed DPT apps. Therefore, the terminology and a classification system may offer some guidance to DPT system operators regarding which measurements to prioritize. DPT developers and operators may also make conscious decisions to integrate measures for epidemic monitoring but should be aware that this introduces a secondary purpose to DPT. Ultimately, the integration of further information (e.g., regarding exact exposure time) into DPT involves a trade-off between data granularity and linkage on the one hand, and privacy on the other. More data may lead to better epidemiological information but may also increase

the privacy risks associated with the system, and thus decrease public DPT acceptance. Decision-makers should be aware of the trade-off and take it into account when planning and developing DPT systems or intending to assess the added value of DPT relative to the existing contact tracing systems.

Keywords: contact tracing app, digital health, COVID-19, coronavirus, performance, effectiveness, digital proximity tracing

INTRODUCTION

Digital proximity tracing (DPT) is a novel health technology, designed to complement manual contact tracing (MCT) by using apps in national efforts to mitigate the Sars-CoV-2 pandemic (1). The primary purpose of DPT is to provide an instrument for fast, anonymous notification of other app users with potential exposure risks to an infected app user (2).

The current Sars-CoV-2 crisis is the first global public health crisis that sees massive, nationwide roll-outs of DPT apps (3). This is noteworthy because DPT has never undergone large-scale, real-world testing in its target population prior to release, as would normally be required for health technologies (4, 5). This is largely due to the urgency of the Sars-CoV-2 crisis, where many countries gave precedence to fast release over extensive testing. All the more, governments face pressure to justify the rapid DPT deployment and to demonstrate its impact on pandemic mitigation. Specifically, evaluations are needed to demonstrate whether single parts and the whole system of DPT perform well from a technical perspective, but also whether DPT helps to contain transmission chains (6). First data-driven studies of DPT effectiveness have started to emerge only recently (7–9).

Ideally, such evaluations of DPT should follow a standardized protocol to allow comparability across countries and settings, but also to facilitate learning from other countries' experiences. However, in direct exchanges and discussions with national health authorities and DPT developers, this group of authors noticed a substantial confusion among health authorities, politicians, and even DPT experts about the aims of DPT, the terminology, and goals for system evaluations. The metaphor of a "Babylonian confusion of tongues" is not too far to describe the current situation. This problem has been recognized, and international health authorities and different groups of academics have attempted to bring some structure into discussions about DPT development, deployment, and evaluation (10, 11). A starting point for discussion (including a glossary with relevant keywords) is presented in von Wyl et al. (6). Furthermore, Colizza et al. present some high-level recommendations regarding effectiveness evaluations of DPT apps (12). Specifically, they emphasize elements such as user update and adherence, speed of notification, but also transparency of DPT risk scoring algorithms and evaluations (12). In addition, experimental studies have been conducted to assess the performance of contact detection by DPT (13–16). The study by Rodriguez and colleagues is noteworthy because it also aimed at defining impact indicators for DPT apps, mostly for user behavior and exposure detection (16). But the study description does not reveal why and how the specific indicators were selected. Of further note, all these studies were conducted by researchers from different scientific backgrounds, and hence using their respective terminologies and performance indicator measures (e.g., from clinical research or computer science), with little consistency across studies.

The present viewpoint attempts to provide further clarifications on key aspects of DPT evaluations by bringing together DPT developers and public health experts from different countries to present a unified proposal for terminology and classification of measures to evaluate DPT. Thereby, we will focus on DPT apps that follow the privacy-preserving design principles outlined by the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) (17) protocol for two reasons. First, the design principles serve—to our knowledge—as the basis for most currently deployed DPT apps [exceptions are, for example, the French TousAntiCovid app (18) or the TraceTogether app from Singapore (13)]. Second, decentralized, privacy-preserving DPT apps, as well as the voluntariness of their use, pose the greatest methodological challenges for monitoring and for designing effectiveness evaluation strategies due to the (intended) paucity of data.

The viewpoint is structured as follows. Section *Principles of Digital Proximity Tracing in Support of Manual Contact Tracing* describes the basic principles of DPT. Section *DPT Is a Complex Intervention* argues that DPT is a complex intervention, relying on the fast completion of clearly defined actions in the notification cascade by different health systems actors. Section *A Closer Look at DPT Steps and Their Influence on Intervention Outcomes* breaks the DPT notification cascade into its separate parts and describes how some basic questions and checks may easily be utilized in the DPT evaluation. Section *Basic Concepts for DPT Evaluations* introduces basic concepts and terminologies to describe and assess DPT systems from different viewpoints, namely system performance assessments and public health effectiveness evaluations. Section *Proposal for Classification of Different DPT Evaluation Measures* outlines a classification matrix to distinguish different types of indicator measures. Concrete indicator examples are provided and referenced in the **Supplementary Materials**. In section *Key Considerations for the Practical Implementation of Performance and Effectiveness Measures*, the viewpoint closes with some basic considerations for developing and implementing indicators for DPT evaluation.

PRINCIPLES OF DIGITAL PROXIMITY TRACING IN SUPPORT OF MANUAL CONTACT TRACING

The principles of DPT have been described extensively elsewhere (1). In brief, DPT enables participants to trace proximity contacts (exposures) that could pose relevant infection risks. If one of the proximity contact persons tests positive for Sars-CoV-2, the app will warn other users who were in close proximity during the infected person's time window of infectivity, thereby increasing the coverage and/or speed of the contact tracing process relative to MCT. The potential advantages of DPT compared with MCT are 3-fold. DPT can (#1) lead to faster exposure notifications than MCT, (#2) reach persons who are not personally known to an index case, and (#3) DPT is easily scalable and should still work when MCT reaches its capacity limits.

Most countries that have released DPT apps have opted for a privacy-preserving, decentralized architecture according to the DP-3T blueprint (17, 19). That is, proximity contacts are not sent to a central server, but stored and evaluated locally on smartphones. The only data that is sent to a central server are pseudonymous, random identifiers of persons with a confirmed SARS-CoV-2 infection. These "infectious" identifiers are downloaded by all other users and compared to the locally stored identifiers to find out which of the proximity encounters were with SARS-CoV-2 positive people. If the temporal aggregation of the matched encounters exceeds a minimal duration at a relevant proximity (depending on the estimated infectiousness of the positively tested person), users are notified and recommendations on further steps to take are provided.

The dominant choice of a privacy-preserving architecture across many countries highlights the emphasis of the primary DPT function ("warning people early in an anonymous manner") over purposes such as disease monitoring. DPT is, first and foremost, a notification tool aimed at breaking transmission chains, and its primary function does not necessitate the collection of personal data of index cases and their contacts. Nevertheless, debates in several countries suggest that DPT is sometimes also viewed as an opportunity to collect data for epidemiological monitoring, for example, to obtain additional information on time and setting of the events with high risk of exposure. Such *secondary functions* of DPT are beyond the scope of this article and are only discussed briefly where relevant for the broader context.

DPT IS A COMPLEX INTERVENTION

The preventive effect of DPT results from a timely warning of exposed persons so that they can enter quarantine and initiate further preventive measures. In this context, timely means a faster contact notification than is usual in MCT. In addition, DPT can also reach persons who would normally be missed by MCT (e.g., because they were chance encounters of the index case). As illustrated by **Figure 1**, the app notification process reflects an information flow in multiple steps to eventually produce specific actions leading to the prevention of further transmission (indicated by the #-signs in **Figure 1**). The effect of DPT depends on the interplay between health system actors (e.g., testing laboratories) and *app users*. It is not the app *per se* but the fast completion of the full notification cascade and subsequent actions that lead to the desired results (9). Of note, the distinction between app users and other actors is warranted because (voluntary) user actions are strongly influenced by behavioral aspects and incentives (20, 21), whereas actions required by other health system actors may depend more on automatization, technical interfaces, resources, or capacity (22).

We identify three high-level tasks (illustrated by colored boxes in **Figure 1**) that combined cover the entire notification cascade:

- **Proximity estimation**: This aspect pertains to the exchange of ephemeral (regularly changing) identifiers between users' devices, and the detection of significant proximity contacts (e.g., $<$1.5 m for more than 15 min).
- **Diagnosis and identifiers upload**: This aspect pertains to the upload of identifiers by index cases.
- **Notification of proximity contacts**: This aspect pertains to the notification of proximity contacts by their mobile device, and the subsequent actions taken by users as a response to this notification.

The dependency of the DPT intervention on its embedding in the overall pandemic mitigation response and the involvement of multiple actors fulfills the definition for complex interventions, as used in other fields of healthcare research (23, 24). In DPT, involved health system actors are setting-specific but may include testing laboratories or health authorities including MCT units or operators of infolines (**Figure 1**). **Figure 2** provides an even more detailed view of 10 individual steps in the DPT notification cascade. The red person illustrates the infected app user who gets tested, receives a positive test result, and triggers the app notification. The green person depicts a proximity contact who receives an app notification. Of note, in most DPT implementations, several of the steps in **Figure 2** involve free user choices whether or not to complete a specific task (e.g., step 6, authorization of key upload) without fears of retribution.

This insight that DPT is a complex intervention involving voluntary actions is relevant from a practical standpoint because it shifts the focus of discussion from single aspects (e.g., technical accuracy of Bluetooth measurements) to a broader systems perspective (25). However, in addition to measuring the final intended outcome of a complex intervention, the monitoring of individual components of a complex intervention is nonetheless important. Because complex interventions, DPT in particular, depend on a seamless, fast cascade of events (as shown in **Figure 1**), measurements characterizing speed and efficiency of specific system components and actors are useful to identify bottlenecks in the notification chain, as well as to act as leverage points for improving system behavior.

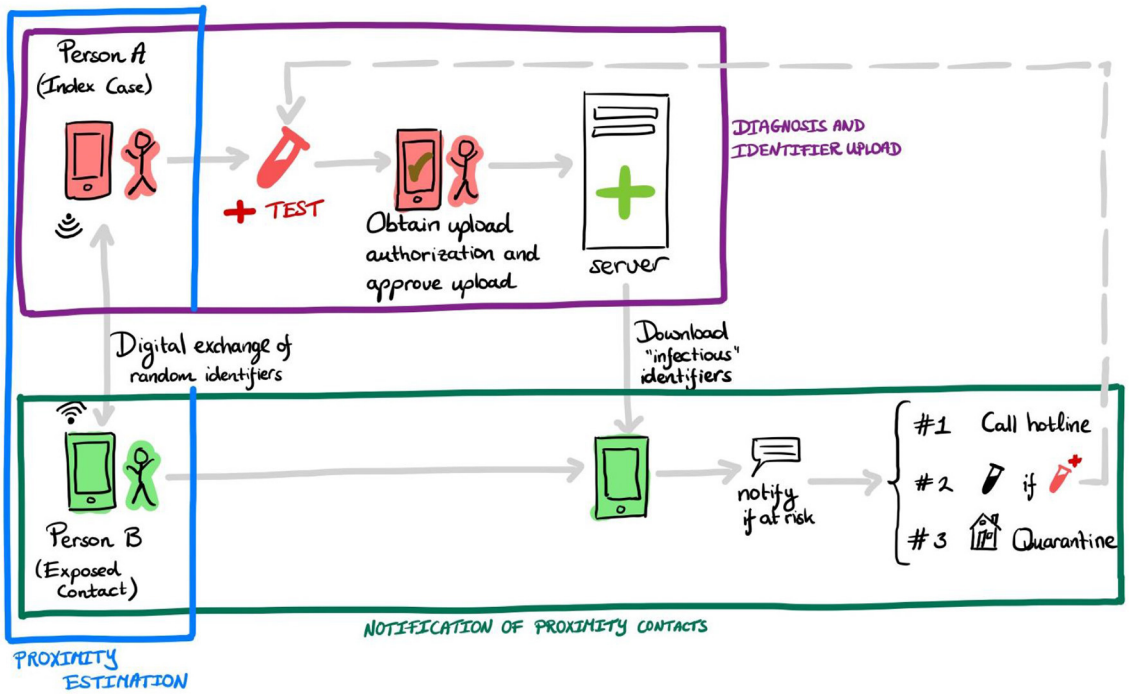


FIGURE 1 | Example of required steps in the notification cascade of digital proximity tracing in decentralized systems. An infected person A receives a positive test result for Sars-CoV-2 and (possibly automatically) an upload authorization. After consent by the user, the app uploads the random identifiers to a server. Person B's device regularly downloads "infectious" identifiers. If B was in close proximity to A (or other infected users) for a prolonged time, then B receives the app notification. Upon receiving this notification, B has several options, for example—in some countries—calling an information line (#1), getting tested for Sars-CoV-2 (#2), or entering quarantine voluntarily (#3). The colored boxes refer to the three main tasks involved in the DPT notification cascade, described in section *DPT Is a Complex Intervention*.

A CLOSER LOOK AT DPT STEPS AND THEIR INFLUENCE ON INTERVENTION OUTCOMES

To illustrate how system components can influence the outcome of the intervention, let us look at the three high-level tasks indicated in **Figure 1** above (*proximity estimation, diagnosis and identifiers upload, proximity contact notification*). The precise details of each of the tasks depend on national or regional choices regarding system design and configuration. However, all systems adhere to a similar, three-step structure. In the following, we will in greater detail describe the processes and country-specific variations, and identify questions that help to assess the system performance.

Proximity Estimation

All systems we consider in this paper rely on the Google/Apple Exposure Notification (GAEN) framework (26). Therefore, the accuracy of proximity estimation depends on the functioning of the GAEN framework and how well the chosen parameters reflect the desired measure of proximity (with slight regional variation).

Determining the accuracy of this component requires accurate ground-truth information about the exact distance and duration of a proximity contact. Collecting such information is nearly impossible in non-experimental settings without infringement of privacy (e.g., because it would require the use of video cameras to establish ground truth). Therefore, research groups (27–30) have used laboratory and simulated settings to replicate these scenarios with carefully constructed/measured ground-truth information. These measurements can then be used to answer questions such as:

- What parameter choices best reflect the desired distance/duration threshold?
- What distribution of false positives/false negatives does this choice induce for particular users' behaviors and under different environments?
- Does the choice of device model/manufacturer/platform influence the realized threshold?

To inform technical questions and setting-specific implementation decisions, the GAEN framework documentation is a suitable starting place (17).

Diagnosis and Identifier Upload

To enable proximity contacts to receive notifications, Sars-CoV-2-positive users must upload the random identifiers they broadcasted during the contagious window. Most countries use a system of upload authorization to allow uploads by confirmed index cases only, thereby preventing false warnings or

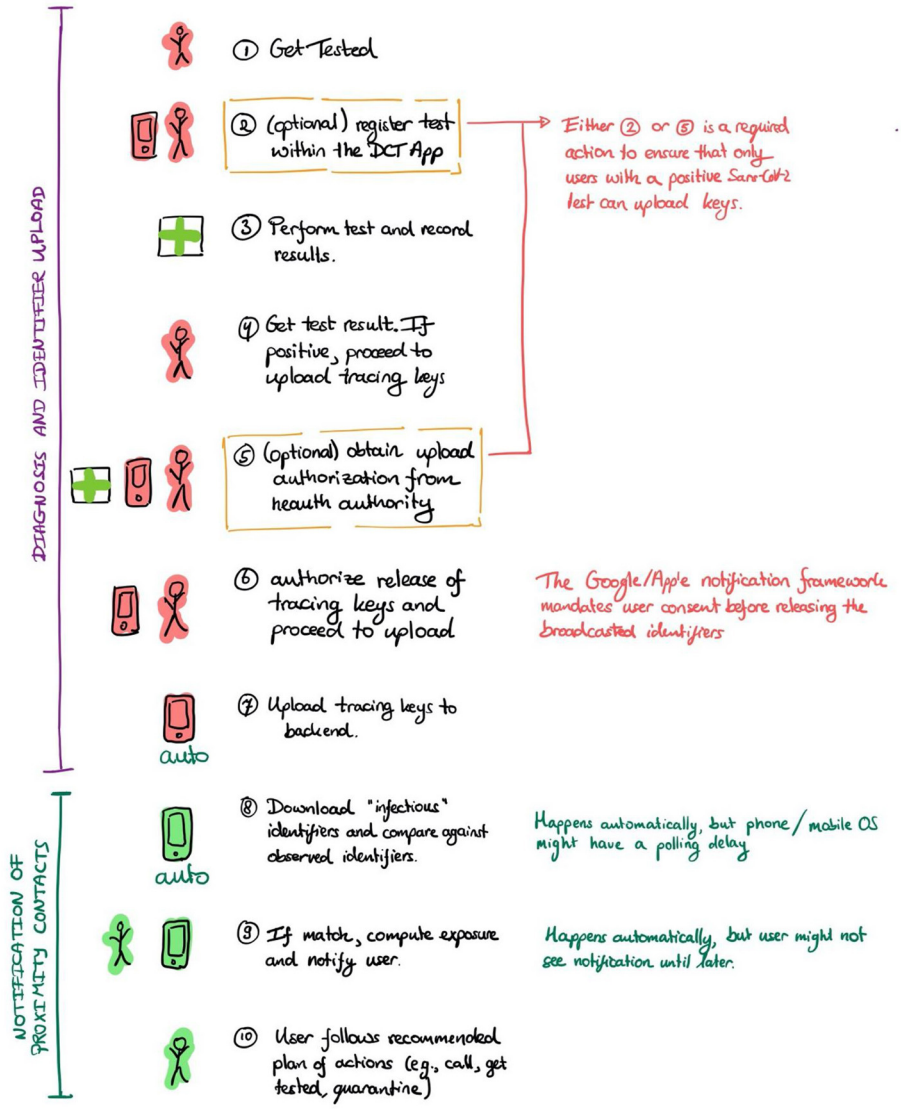


FIGURE 2 | Specific actions required in the notification cascade.

manipulation. As of today, automatic uploads of identifiers are not possible.

The following steps are considered critical (**Figure 2**):

- Get tested and receive a (positive) test result.
- Obtain upload authorization by a health authority (e.g., automatically when users registered the test in the app; or via an upload authorization code obtained through interaction with the health authorities after receiving a positive test).
- Upload identifiers by consenting to the release and upload of identifiers in the app.

The exact process of obtaining upload authorizations differs per country. Some countries, e.g., Belgium (31), and Germany (32), let users register Sars-CoV-2 tests in the app (who then also receive their test result through the app). Users with positive tests then automatically receive an upload authorization. In addition to or instead of such an automatic flow based on registered tests, other countries provide index cases with an upload authorization code that they enter into the app (Switzerland and Portugal use this as the only flow, German and Belgium use it as an alternative) or let users obtain authorizations via an eHealth system (e.g., in Estonia where the eHealth system also manages Sars-CoV-2 tests).

Moreover, the GAEN framework requires users to explicitly consent to release the random identifiers their phones have broadcast. Therefore, any app will request user action and explicit consent. However, the timing of consent provision varies across countries and can occur, for example, during laboratory test registration (e.g., in Germany) or upon receipt of a positive test result.

Based on the three critical processes described above, one could ask the following questions related to the operation of the system:

- How long between users getting tested and receiving a result?
- What fraction of users register tests in the app, if this function is available?
- What is the time from receiving the result to uploading identifiers?
- How many users with permissions to upload do not finish the process? At what point do they abandon the process?

While data collections to address these questions are highly setting-specific, publicly accessible examples of such monitoring systems already exist, for example, in Switzerland (9, 33), Germany (34), or in the Netherlands (35).

Notification of Proximity Contacts

Apps regularly retrieve uploaded positive identifiers from the system's backend server and then compare them (using the GAEN framework) against stored identifiers. If the framework detects a sufficiently long exposure, the app sends a notification to the end-user. Depending on the country, this notification instructs the user to contact the local health authority or a hotline, take a test, or self-quarantine. As with MCT, users may fail to follow these instructions.

To protect privacy (and to facilitate efficiency), uploaded identifiers are not immediately downloadable by other app end-users. Instead, the backend regularly releases a new batch of identifiers. For example, every 2 h.

The following factors influence the notification timeline.

- Time to publish. The time between receiving "positive" identifiers at the system's backend, and their publication, when these identifiers can be retrieved by other phones.
- Polling frequency. The frequency with which user's devices retrieve new "positive" identifiers and compute matches. This time is influenced by the mobile OS's scheduler as well as the internet connectivity of the phone.
- Time for the user to notice the notification and to subsequently act upon it.

The first and second factors are configurable parameters, constrained by the capabilities enabled by the GAEN framework (17). The effect of the system's scheduler can be tested in a laboratory setting.

The following questions can be asked concerning this notification task (some may not be possible to answer in privacy-preserving architectures):

- How many people receive a notification? How many of them later test positive?
- How many people follow through after receiving a notification?
- What is the time between positive upload and notification?
- How long between a notification is received and the user acts upon it?

The monitoring websites mentioned in section *Diagnosis and Identifier Upload* also provide good example metrics for the proximity contact notification step (33–35).

BASIC CONCEPTS FOR DPT EVALUATIONS

As illustrated above (**Figures 1**, **2**), DPT relies on a fast and seamless flow of information along the notification cascade. Blockades or inefficiencies in single steps can lead to bottlenecks and prevent the information flow, thus inhibiting the primary goal of DPT to warn other app users about potential risk exposures. The information flow of each DPT step can be described by at least three attributes: *speed, yield, and capacity*. *Speed* describes how fast an action is completed and can be measured in terms of time. *Yield* refers to a completion fraction, that is, a number of tasks executed as needed per 100. Yield sometimes also has a time connotation, that is the fraction of task completion for a given time frame (also referred to as *throughput*). *Capacity* relates to the task volume an actor can handle in a given time (which, in turn, may also influence speed and yield). For example, testing laboratories or manual contact tracers can only process a certain number of samples or cases, given the available resources such as machinery or personnel. Therefore, volumes can be described as percentages below or above the capacity limit.

The different DPT system components, as well as the basic attributes (*speed, yield, capacity*) and questions about the DPT functioning, lend themselves to the development of *Key Performance Indicators* (KPIs). For a better interpretation of these KPIs, a contextualization with the dynamic of the Sars-CoV-2 pandemic is often useful (e.g., to compare the number of key uploads with the incidence of new infections).

Overall, such KPIs provide valuable information on the procedural performance of the overall system, as well as possible bottlenecks in the notification cascade. If any of the components of the tasks in **Figure 1** (e.g., providing upload authorization) malfunctions, the delay will ripple to the whole notification cascade and undermine DPT's ability to perform its primary purpose of notifying exposed contacts. Therefore, KPIs can also be viewed as measures of technical and procedural preconditions for DPT to fulfill its primary purpose (9).

However, KPIs have the caveat that they often summarily reflect sequences of different actions (e.g., app usage, positive test, identifier uploads, download of identifiers, and proximity estimation). This composition complexity hinders the interpretation of these metrics as a consequence of a single factor.

Furthermore, KPIs are, in a strict sense, not revealing concerning how well DPT achieves its primary purpose of reducing viral transmissions, respectively its "*effectiveness,*" defined as the "ability to produce a desired result" (36). In epidemiological studies, the concept of effectiveness often stands for the real-world effect of an intervention against a comparator (comparative effectiveness) and is expressed as an exposure-outcome relationship.

It is important to note that, in the majority of countries, DPT is intended and designed to complement MCT. Therefore, in most settings, a "fair" DPT effectiveness evaluation would involve comparison between the use of DPT apps vs. non-use in addition to manual contact tracing. In those settings where DPT is implemented alongside MCT, effectiveness investigations should ideally center around the three postulated main advantages of DPT over MCT: DPT should lead to faster exposure notifications than MCT, DPT can reach persons who are not personally known to an index case, and DPT should still work when MCT reaches its resource limits (6). Specific effectiveness outcomes could focus, for example, on the time from the first exposure notification (either by MCT and/or DPT) to entering quarantine, or on the comparisons of the average number of persons who later test positive between groups who were notified by MCT or DPT. In both examples, the most obvious comparator is classic MCT. In settings where no MCT exists (or where it is no longer operable) and DPT is introduced in a staggered process, regional comparisons of Sars-CoV-2 incidence could be performed between geographic units that have introduced DPT at different times.

Sometimes DPT and DPT-related measurements are also discussed in the context of *epidemiological monitoring* (surveillance). Such discussions are tied to the hope of gathering relevant data and gaining insights about transmission dynamics. It is important to note that epidemiological monitoring is not part of the key functionality of DPT and necessitates an entirely different set of measurements and KPIs that go beyond the data requirements for privacy-preserving proximity tracing and it is not included in the following discussions.

PROPOSAL FOR CLASSIFICATION OF DIFFERENT DPT EVALUATION MEASURES

High-Level Distinction Between Key Performance Indicators and Public Health Effectiveness Metrics

Figure 3 proposes a *KPIs classification matrix* of different measure types and perspectives ("aspects") relevant for the DPT assessments. We acknowledge that the distinction between the different proposed types may not always be clean-cut in practice. In fact, one may need several of these metrics to assess how well the system performs each of the three tasks DPT systems must realize to fulfill their objective (see **Figure 1**). Nevertheless, the classification matrix may guide KPI development by illustrating different dimensions that comprehensive DPT monitoring systems should cover.

The horizontal dimensions of the classification matrix show different steps from a basic Input-Processing-Output (IPO) model perspective (37). Each step in the DPT notification cascade requires resources (e.g., technical infrastructure, money) and inputs (e.g., information), which are processed to create outputs (e.g., notifications). The different IPO steps can be examined from different viewpoints shown in the vertical classification matrix dimension, namely from a technical (app-) perspective, as well as from the viewpoint of different actors, including app end-users, but also laboratories or public health services (**Figure 1**).

Therefore, each matrix cell reflects a combination of IPO-step and viewpoint that can be useful to describe and evaluate the performance of specific steps or elements of DPT systems (KPIs). By taking a specific step in the notification cascade described in section *A Closer Look at DPT Steps and Their Influence on Intervention Outcomes*, the performance can be evaluated from different angles using the guiding questions such as: What are the resources needed to complete this step? How well does the information flow along the notification cascade work? Or how much desired output is generated by this step? Such KPIs can be formalized as raw numbers, proportions, or ratios to describe *speed, yield, and capacity* attributes (as described in section *Basic Concepts for DPT Evaluations*).

Separate and located below the classification matrix in **Figure 3** are the public health *effectiveness* measures. They are distinct from KPIs and aim to address a different question: does the DPT system achieve its intended primary aim of notifying exposed app users swiftly so they can take preventive measures? Measures of the DPT effectiveness can relate, for example, to the prevention of further transmission or comparative cost-effectiveness when compared to MCT (6). The health-centered effectiveness measures are different from KPIs, and yet not independent. The completion of the notification process is a precondition for achieving the DPT public health goal. In other words, many public health metrics are an integral of different processes in the app notification cascade, as they are a direct consequence of how effective the notification cascade tasks are executed.

A Worked Example of the KPIs Classification Matrix and Public Health Measures: The Proximity Notification Step

To further illustrate the use and usability of the KPIs classification matrix, we will—cell by cell—describe KPIs examples related to the proximity notification step (which could also be applied to classic MCT). The proximity notification step (section *Notification of Proximity Contacts*) is a crucial element in the notification cascade with a direct relation to the primary DPT goal: to warn proximity contacts as early as possible about potential transmission risk exposures.

The first matrix column represents resource-oriented metrics, which define resource needs for technical and non-technical implementation and include, for example, costs for PCR-tests (which are free for persons with a DPT notification in some countries), costs for quarantining of DPT-notified persons, or any other expenditures.

In the vertical dimension, the first matrix row reflects the app-oriented, technical perspective. During DPT development and operation, IT system design requires choices for parametrization of measurements and backend systems, which are resource-dependent and may impact speed, yield, or capacity of the DPT processes. Therefore, concrete KPIs examples for resource-oriented metrics from the technical perspective (*cell 1*) include costs for development and maintenance of the app itself, as

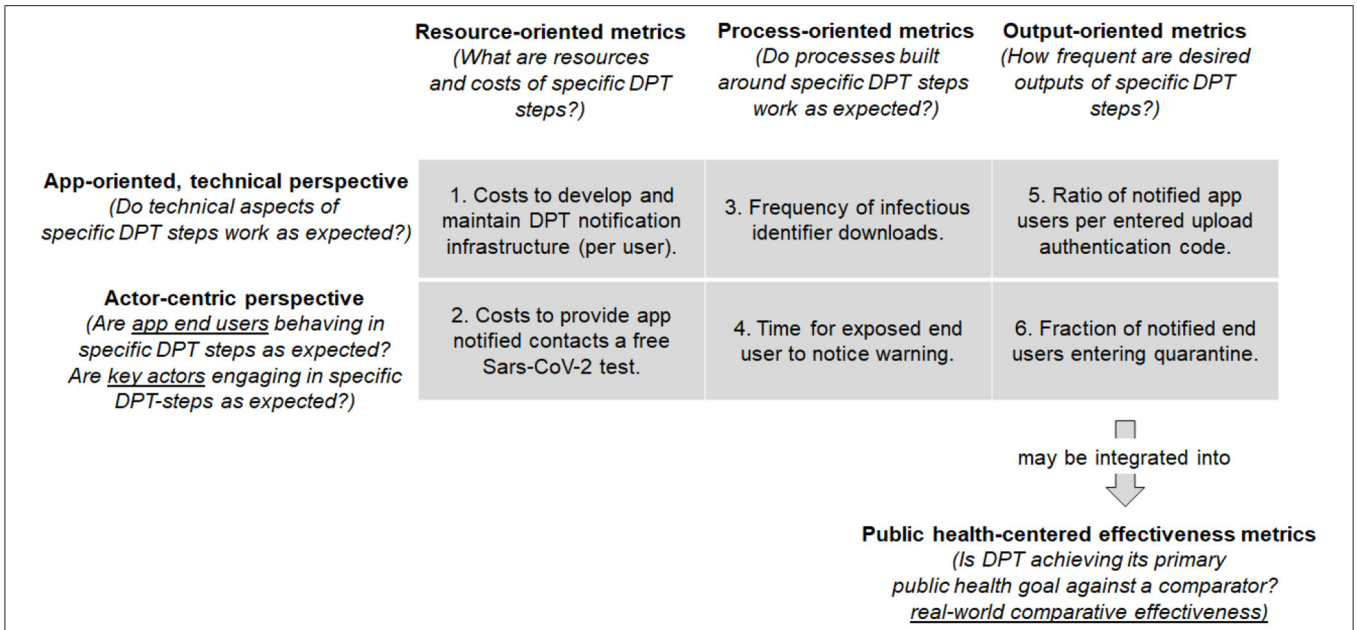


FIGURE 3 | Proposal for classification matrix of different indicator types to monitor DPT systems by steps of the Input-Process-Output model (resource-, process-, and output metrics) as well as by aspect (app-oriented, technical viewpoint vs. actor-centric perspectives). The examples in the matrix pertain to the notification step of proximity contacts (section *Notification of Proximity Contacts*). The output oriented metrics (5 & 6) may be integrated into public health-centered metrics.

well as for the technical infrastructure and the backend. Such expenditures are often scaled by the number of active users or the number of quarantine orders (or other indicators for prevented transmissions).

The second matrix row reflects the actor-centric perspective on the notification chain. Involved actors include laboratories that perform PCR-tests and communicate results to app end-users and health authorities, health authorities and other authorized parties that take calls from notified users, and—in settings with manual upload authorization—the parties who provide upload authorization codes to diagnosed users. Furthermore, among all involved actors, the app end-users play a central role. End-users need to decide whether to use the DPT app, but also to actively trigger (or at least consent to) the upload of identifiers in case of testing positive for Sars-CoV-2. Example KPIs that combine resource and actor perspectives (*cell 2*) include expenditures for user-linked actions, such as the costs for providing app-notified contacts a free Sars-CoV-2 test.

Process metrics are located in the second matrix column. Those KPIs describe interactions of the app and its users with other parts and actors of the health system. For the app to work as intended, several processes need to occur seamlessly so that all tasks can be carried out successfully and timely: from testing to prompt results communication, upload authorization code generation, and identifier upload, notification of exposed contacts, and these contacts taking action (e.g., calling the hotline or a doctor and receiving advice). Process metrics can be used to monitor how well the different conditions for app-functioning are met, respectively, whether the different system parts work as expected.

Examples of process-centric metrics that integrate the technical perspective (*cell 3*) include, for example, precision and recall of Bluetooth and exposure time measurements, which are usually assessed in experimental settings. Specific design choice evaluations may involve measurements of how well the GAEN/Bluetooth approximation reflects actual physical distance and time exposure, as well as backend configurations regarding the frequency of infectious key uploads or downloads of lists of infectious identifiers (which only happens a certain number of times per day). *Cell 4* represents process-oriented metrics from an actor perspective. Several steps in the notification cascade require human involvement, sometimes on a voluntary basis. Therefore, such actions may be strongly influenced by behavioral aspects, digital and health literacy, but also by incentives. KPIs examples include the fraction of positive app users who consent to or actively initiate the identifier uploads. KPIs used to describe such steps can often be based on yield (fraction of completed tasks) or speed (time to completion) attributes.

Finally, the third matrix column reflects output-oriented metrics. These KPI refer to desired outputs of DPT, which could be numbers or yields of DPT-notified users who undertake a recommended action (e.g., entering quarantine or getting tested for Sars-CoV-2). These metrics are related to public health goals but differ in that they focus on an intra-system perspective: they often do not encompass external comparators but focus on how a system has evolved. Technical aspects influence desired outcomes in various ways. In *cell 5*, notifications of exposed contacts are the desired outcome. An example KPI is the ratio of the number of exposure notifications over the number of upload authentication codes entered by positive tested users. The

ratio depends on various technical aspects, including calculation methods for proximity risk scores derived from Bluetooth attenuation measurements.

Furthermore, many outputs depend on end-user interactions. App-notified contacts are expected to follow certain procedures such as calling an infoline, getting tested, or entering quarantine (*cell 6*). Examples for such user-dependent outcome KPIs are the fraction app-notified users who voluntarily enter quarantine or who seek testing. It is often instructive to express these KPIs both as yield (fraction who completes an action) and speed attributes (time until an action is completed).

Finally, the public-health-oriented metrics reflect the real-world effectiveness of DPT apps as defined above and relate to the main pandemic mitigation goals. These DPT goals are the result of the interplay between the different technical and non-technical aspects of the notification cascade. That is, the KPIs provide information about the performance of the notification cascade required for DPT effectiveness. However, KPIs do not provide direct evidence for DPT effectiveness, for which a comparator group is required (which, however, can also be a comparator without any measures such as DPT or MCT as described in section *Basic Concepts for DPT Evaluations*). In the context of the proximity contact notification step, an effectiveness evaluation could, for example, compare the time from positive testing of the index case to quarantine of the proximity contact between DPT and standard MCT.

As outlined in the previous section, key performance indicators may contribute to effectiveness evaluations if (a) they reflect actions or outcomes that are of public health relevance and (b) if they are compared against a suitable "comparator group." Condition (b) is also the reason why the indicator in cell 6 is not (yet) an effectiveness measure. Monitoring this indicator in a time-series will be informative about longitudinal changes. However, an effectiveness analysis would require to compare the indicator 6 against a reference group, such as the fraction of persons entering voluntary quarantine with or without exposure notification. The latter can occur if exposed contacts receive an informal exposure notification, e.g., a message from a friend or a relative who tested positive for SARS-CoV-2.

KEY CONSIDERATIONS FOR THE PRACTICAL IMPLEMENTATION OF PERFORMANCE AND EFFECTIVENESS MEASURES

Definition of Expected Process Targets

To monitor the process performance of DPT, it is helpful to have an expectation of where indicators should stand at a given time. That is, to assess performance on the basis of absolute numbers (e.g., number of authorized uploads) or a yield (e.g., fraction of authorized over realized uploads), expectations or precise benchmark targets should be defined. Because DPT is still a very novel health technology, defining specific benchmarks can be challenging. Moreover, targets may not only be country—but also setting—or subgroup-specific (e.g., targeting a specific app coverage in the working population).

Target definitions can in part be informed by modeling results (e.g., concerning required DPT coverage to create an effect) (1, 38–40). But in many instances, only qualitative targets may be feasible because of a lack of suitable reference values. A possible approach to derive such qualitative targets is to describe "desired" effects in a hypothetical, perfectly functioning system. For example, KPIs to measure the full completion of actions, such as the fraction of positive users that upload their information to the server (e.g., by measuring the fraction of upload authorization codes that are redeemed). Other KPI measuring attributes such as speed may lack a clearly defined benchmark (for example, the time from app notification to quarantine), but could be formulated in terms of a comparison to other measures (for example, the same intervals in a manual process).

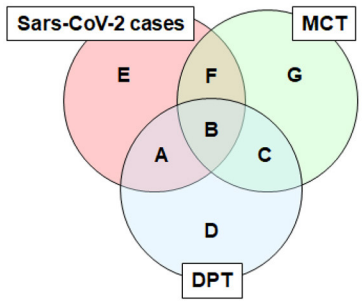
Practical Implications of the Distinction Between Performance and Effectiveness Measures

The distinction between KPIs and comparative effectiveness measures is more than just semantics. KPIs and effectiveness measures require different data and measurement approaches. Process metrics can be collected at different contact points in the notification cascade (**Figure 1**), for example at app download, during regular configuration updates, when upload authorization codes are generated, or when notified users call an infoline. But due to the privacy-preserving, decentralized nature of DPT apps (at least those that follow the DP-3T blueprint) these metrics provide only aggregated, non-identifiable data, and these data points cannot be easily connected into a unique data stream for a specific user.

By contrast, comparative effectiveness investigations need to establish a link to processes and data collections of its comparator, which will be MCT in most settings. But from a privacy perspective, it should be clear that investigations providing unquestionable evidence for DPT effectiveness can no longer be privacy-preserving and anonymous. For example, an ideal study of DPT/MCT effectiveness should be able to connect infected cases with exposed contacts and follow their notification and quarantining cascade in great detail. Often proposed key metric is the secondary attack rate (SAR), which measures how many exposed contacts of an infected person later test positive for Sars-CoV-2 (11). However, calculating precise SAR measures require an exact identification and linkage of cases and contacts, something that is not foreseen in privacy-preserving DPT apps. At the same time, DPT can also mitigate some shortcomings of MCT, for instance by improving notification speed and extending to exposed contacts who are unknown to the index case.

Choosing the Right Denominator

Selecting appropriate denominators for KPIs and public health metrics can pose challenges. While there is no universal best practice, we find that Venn diagrams can be a helpful tool to guide the search for suitable denominators (**Figure 4**). Venn diagrams are useful to illustrate the (non-)overlap between different populations of interest. In the context of DPT, these are the persons who are tested positive for Sars-Cov-2 (cases), those who



Example interpretation

- All persons with a positive Sars-CoV-2 test (cases): Segments A + B + F + E
- All persons who are contacted by manual contact tracing. Infected cases: Segments B + F Exposed contacts: Segments C + G
- All persons who received a DPT app notification: Infected cases: Segments A + B Exposed contacts: Segments D + C

The overlapping segments A, B, C, F indicate groups that belong to more than one group.

The interpretation of segments can be adjusted as needed, but the assumptions of the Venn diagram need to be clarified along the criteria listed in the main text.

FIGURE 4 | Example use of Venn diagrams to define denominators for KPIs and public health metrics (courtesy of https://pact.mit.edu/).

are notified by DPT, and those who are informed through manual contact tracing (MCT). DPT represents the whole population of notified app users. MCT includes all persons who were identified through manual contact tracing. The segments labeled from A-G represent population counts (e.g., corresponding to outcome-oriented metrics in **Figure 3**).

While Venn diagrams may facilitate conceptual discussions, the assumptions and context must be well-described. These assumptions include the following aspects.

- Goal: What type of high-level task or metric should be described by the Venn diagram?
- Population: What is the origin of the data for KPIs or public health metric analysis (case series, cohort study, population-based study, or administrative database)?
- Time horizon: What is the time perspective covered by the Venn diagram (cross-sectional, or cumulative over a longer period)?
- Evaluation time point: At what time point are classifications into the three groups (positive tested cases, DPT, MCT) established? Shortly after the time of exposure, when PCR test results are still pending? Cross-sectional at a given moment in time?
- Case definitions: Furthermore, what is the accuracy with which infection status can be determined (i.e. how to deal with infected, untested individuals)?
- Setting-specific assumptions: Finally, country-specific simplifications may be warranted based on the Test-Trace-Isolate-Quarantine strategy (e.g., whether all PCR-positive cases are automatically referred to MCT). Therefore, for some countries, one can assume that segments A + E are close to 0, whereas F + B are approximating the number of positive cases.

Feasibility of Integrating Measurements Directly Into DPT

Given the different data requirements for KPI monitoring and effectiveness studies, the question arises how the necessary information should be collected: by integration into DPT apps and corresponding backend systems or through separate research studies?

The addition of measurement capabilities to DPT apps can be a sensitive matter. First, DPT apps following the DP-3T blueprint are not designed as data collection instruments, but as privacy-preserving notification tools that keep their users anonymous. Adding more measurement capabilities (e.g., in the backend or the app itself) leads to a data granularity-privacy trade-off. The gain in knowledge has to be weighed against a greater likelihood for de-anonymization. Adding measurement capabilities may require an increased trust by end-users in the system operators. For example, collecting exact dates of exposures, notifications, and contacts with different actors (e.g., the infoline) may, in combination, imply that study subjects may no longer be sure that their identities remain concealed. The combination of these measures may already identify persons uniquely, especially in smaller populations. If the collection of such data is to take place, it must happen transparently and app users should provide informed consent. In addition, other privacy-preserving technologies that minimize the amount of data collected and limit the capability of linkage across databases can be employed[1].

The decision of whether and how to integrate additional measurements into DPT apps (beyond what is needed for notification) is one that each country needs to make separately. Such a decision must take into account specific legal considerations, overall acceptance of the DPT technology, and public expectations toward DPT privacy, as well as the individual and societal risks associated with the new data collection.

As an alternative to DPT-integrated measurements, dedicated (observational) research studies with volunteers and specifically designed databases should be considered. Given informed consent by participants, a linkage of information between DPT and MCT should be possible. For example, studies could be

[1]Examples of privacy-preserving building blocks that could be used to support measurements while minimizing risks for users are multi-party computation, differential privacy, anonymous authentication, and homomorphic encryption. Other privacy technologies could also be of interest.

integrated into contact tracing and specifically survey index cases and exposed contacts regarding app usages and notifications. Ireland, for example, employs separate case management systems, which collect numerous complementary data that could be valuable for effectiveness research. Alternatively, app users could be presented with short questionnaires including questions regarding usage and past exposures. However, a linkage of apps with survey advertisements (even on a voluntary basis) could be regarded as intrusive and fuel privacy fears. Therefore, the advantages and disadvantages of each survey recruitment method should be deliberated carefully.

CONCLUSION

The development of monitoring systems for DPT performance and effectiveness requires complex decisions. While there is no universal advice that could suit all settings and countries, it may help to obtain clarity on the distinctions between performance monitoring and effectiveness. Furthermore, decision makers should become aware that *not all measurements can and should be integrated into DPT apps* and connected backend systems. Separate studies or data collection systems may be needed to generate the necessary evidence for performance and effectiveness of DPT. The proposed indicator classification aims to support this process.

AUTHOR CONTRIBUTIONS

VW conceived the study and wrote the first draft. All authors provided conceptual inputs, analyzed the data, contributed to paper writing and revision, and approved the manuscript.

ACKNOWLEDGMENTS

The authors thank the regular participants of the DP-3T international exchange group for helpful discussions. VW thanks the MIT PACT group members https://pact.mit.edu/ for sharing their insights and for providing the inspiration to use Venn diagrams. We also thank Mina Stanikić for language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.677929/full#supplementary-material

REFERENCES

1. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. (2020). 368:6491. doi: 10.1126/science.abb6936
2. Swiss National Covid-19 Science Task Force. *Digital Proximity Tracing. Policy Briefs*. (2020). Available online at: https://ncs-tf.ch/de/policy-briefs/digital-proximity-tracing-15-may-20-en/download (accessed July 19, 2021).
3. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. *Comput Sci Rev*. (2020) 38:100307. doi: 10.1016/j.cosrev.2020.100307
4. Servick K. COVID-19 contact tracing apps are coming to a phone near you. How will we know whether they work? *Science*. (2020). doi: 10.1126/science.abc9379
5. Zastrow M. Coronavirus contact-tracing apps: can they slow the spread of COVID-19? *Nature*. (2020). doi: 10.1038/d41586-020-01514-2
6. von Wyl V, Bonhoeffer S, Bugnion E, Puhan MA, Salathé M, Stadler T, et al. A research agenda for digital proximity tracing apps. *Swiss Med Wkly*. (2020) 150:w20324. doi: 10.4414/smw.2020.20324
7. Menges D, Aschmann HE, Moser A, Althaus CL, von Wyl V. A data-driven simulation of the exposure notification cascade for digital contact tracing of SARS-CoV-2 in Zurich, Switzerland. *JAMA Netw Open*. (2021) 4:e218184. doi: 10.1001/jamanetworkopen.2021.8184
8. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. *Nature*. (2021) 594:408–12. doi: 10.1038/s41586-021-03606-z
9. Salathé M, Althaus CL, Anderegg N, Antonioli D, Ballouz T, Bugnion E, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. *Swiss Med Wkly*. (2020) 150:w20457. doi: 10.4414/smw.2020.20457
10. World Health Organization & European Centre for Disease Prevention and Control. *Indicator Framework for the Evaluation of the Public Health Effectiveness of Digital Proximity Tracing Solutions*. World Health Organization (2021). Available online at: https://apps.who.int/iris/handle/10665/341818 (accessed July 19, 2021).
11. European Centre for Disease Prevention and Control. *Mobile Applications in Support of Contact Tracing for COVID-19 - A Guidance for EU EEA Member States*. (2020). Available online at: https://www.ecdc.europa.eu/en/publications-data/covid-19-mobile-applications-support-contact-tracing#no-link (accessed July 19, 2021).
12. Colizza V, Grill E, Mikolajczyk R, Cattuto C, Kucharski A, Riley S, et al. Time to evaluate COVID-19 contact-tracing apps. *Nat Med*. (2021) 27:361–2. doi: 10.1038/s41591-021-01236-6
13. Huang Z, Guo H, Lee Y-M, Ho EC, Ang H, Chow A. Performance of digital contact tracing tools for COVID-19 response in Singapore: cross-sectional study. *JMIR Mhealth Uhealth*. (2020) 8:e23148. doi: 10.2196/23148
14. Sattler F, Ma J, Wagner P, Neumann D, Wenzel M, Schäfer R, et al. Risk estimation of SARS-CoV-2 transmission from bluetooth low energy measurements. *NPJ Digit Med*. (2020) 3:129. doi: 10.1038/s41746-020-00340-0
15. Leith DJ, Farrell S. Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a light-rail tram. *PLoS ONE*. (2020) 15:e0239943. doi: 10.1371/journal.pone.0239943
16. Rodríguez P, Graña S, Alvarez-León EE, Battaglini M, Darias FJ, Hernán MA, et al. A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. *Nat Commun*. (2021) 12:587. doi: 10.1038/s41467-020-20817-6
17. Troncoso C, Payer M, Hubaux J-P, Salathé M, Larus J, Bugnion E, et al. *Decentralized Privacy-Preserving Proximity Tracing*. (2020). Available online at: https://arxiv.org/abs/2005.12273 (accessed July 19, 2021).
18. TousAntiCovid. Available online at: https://bonjour.tousanticovid.gouv.fr/ (accessed April 20, 2021).
19. Wikipedia contributors. *COVID-19 Apps Wikipedia, The Free Encyclopedia*. (2020). Available online at: https://en.wikipedia.org/w/index.php?title=COVID-19_apps&oldid=994533835 (accessed December 22, 2020).
20. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in switzerland: panel survey analysis. *JMIR Public Health Surveill*. (2021) 7:e25701. doi: 10.2196/25701
21. Bonardi J-P, Brülhart M, Danthine J-P, Saxena A, Thöni C, Thoenig M, et al. How to make digital proximity tracing work: the view from

economics. (2020). Available online at: https://e4s.center/document/how-to-make-digital-proximity-tracing-work-the-view-from-economics/ (accessed July 19, 2021).
22. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert J, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health*. (2020) 5:e452–9. doi: 10.1016/S2468-2667(20)30157-2
23. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med*. (2010) 8:63. doi: 10.1186/1741-7015-8-63
24. Ross J, Stevenson F, Dack C, Pal K, May C, Michie S, et al. Developing an implementation strategy for a digital health intervention: an example in routine healthcare. *BMC Health Serv Res*. (2018) 18:794. doi: 10.1186/s12913-018-3615-7
25. von Wyl V. Challenges for nontechnical implementation of digital proximity tracing during the COVID-19 pandemic: media analysis of the SwissCovid App. *JMIR Mhealth Uhealth*. (2021) 9:e25345. doi: 10.2196/25345
26. Apple/Google. *Privacy-Preserving Contact Tracing*. (2020). Available online at: https://www.apple.com/covid19/contacttracing (accessed July 19, 2021).
27. bt-measurements. Github. (2020). Available online at: https://github.com/DP-3T/bt-measurements (accessed December 22, 2020).
28. PT-System-Documents. Github. (2020). Available online at: https://github.com/admin-ch/PT-System-Documents (accessed December 22, 2020).
29. cwa-documentation. Github. (2020). Available online at: https://github.com/corona-warn-app/cwa-documentation (accessed December 22, 2020).
30. gaen-risk-scoring. Github. (2020). Available online at: https://github.com/lfph/gaen-risk-scoring (accessed December 23, 2020).
31. Sciensano Belgian institute for health. *Coronalert be Frequently Asked Questions: How do I Report an Infection?* Brussels: Sciensano (2020).
32. Hoerdt J. *What Requirements Must Be Met in Order for a Personal Test Result to Be Successfully Transmitted via the Corona-Warn-App?*. Corona-Warn-App open-source project (2020). Available online at: https://www.coronawarn.app/en/blog/2020-09-11-positive-test-result/ (accessed December 18, 2020).
33. Swiss Federal Office of Statistics. *SwissCovid App Monitoring*. (2020). Available online at: https://www.experimental.bfs.admin.ch/expstat/de/home/innovative-methoden/swisscovid-app-monitoring.html (accessed July 19, 2021).
34. RKI - Coronavirus SARS-CoV-2 - *Übersicht zu aktuellen und früherer Zahlen und Fakten zur Corona-Warn-App*. (2020). Available online at: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/WarnApp_KennzahlenTab.html (accessed December 23, 2020).
35. Onderzoek: hoe weten we of CoronaMelder helpt tegen Corona? (2020). Available online at: https://www.coronamelder.nl/nl/faq/40-onderzoek-hoe-weten-we-of-coronamelder-helpt-tegen-corona/ (accessed December 2 3, 2020).
36. Definition of EFFECTIVENESS. Available online at: https://www.merriam-webster.com/dictionary/effectiveness (accessed December 2 2, 2020).
37. Wikipedia contributors. IPO model. Wikipedia, The Free Encyclopedia. (2020). Available online at: https://en.wikipedia.org/w/index.php?title=IPO_model&oldid=974000798 (accessed December 22, 2020).
38. Abueg M, Hinch R, Wu N. Liu L, Probert WJM, Wu A, et al. Modeling the effect of exposure notification and non-pharmaceutical interventions on COVID-19 transmission in Washington state. *npj Digit Med*. (2021) 4:49. doi: 10.1038/s41746-021-00422-7
39. Cencetti G, Santin G, Longa A, Pigani E, Barrat A, Cattuto C, et al. Digital proximity tracing on empirical contact networks for pandemic control. *Nat Commun*. (2021) 12:1655. doi: 10.1038/s41467-021-21809-w
40. López JAM, Arregui-Garcla B, Bentkowski P, Bioglio L, Pinotti F, Boëlle PY, et al. Anatomy of digital contact tracing: Role of age, transmission setting, adoption, and case detection. *Sci Adv*. (2021) 7:eabd8750. doi: 10.1126/sciadv.abd8750

Conflict of Interest: DB is an employee of Cybernetica AS.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lueks, Benzler, Bogdanov, Kirchner, Lucas, Oliveira, Preneel, Salathé, Troncoso and von Wyl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.