Rapid Open Development and Clinical Validation of Multiple New 3D-Printed Nasopharyngeal Swabs in Response to the COVID-19 Pandemic

Cody J Callahan\textsuperscript{a}, Rose Lee\textsuperscript{bc}, Katelyn E. Zulauf\textsuperscript{bd}, Lauren Tamburello\textsuperscript{a}, Kenneth P. Smith\textsuperscript{bd}, Joe Previtera\textsuperscript{f}, Annie Cheng\textsuperscript{b}, Alex Green\textsuperscript{bd}, Ahmed Abdul Azim\textsuperscript{c}, Amanda Yano\textsuperscript{g}, Nancy Doraisswami\textsuperscript{h}, James E. Kirby\textsuperscript{bd}, and Ramy Arnaout\textsuperscript{bdj}

\textsuperscript{a}Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{b}Clinical Microbiology Laboratories, Division of Clinical Pathology, Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA USA 02215.
\hline
\textsuperscript{c}Division of Infectious Disease, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{d}Harvard Medical School, Boston, MA, USA 02115.
\hline
\textsuperscript{e}Division of Urologic Surgery, Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{f}Division of Respiratory Therapy, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{g}Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{h}Division of Perioperative Services, Department of Central Processing, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{i}Division of Infection Control/Hospital Epidemiology, Silverman Institute for Healthcare Quality and Safety, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215.
\hline
\textsuperscript{j}Division of Clinical Informatics, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA USA 02215.
\hline
\textsuperscript{†}To whom correspondence should be addressed at rarnaout@bidmc.harvard.edu
ABSTRACT

Background
The SARS-CoV-2 pandemic has caused a severe international shortage of the nasopharyngeal (NP) swabs required for testing. We participated in an unprecedentedly open and urgent iterative process to help develop and validate new swabs.

Methods
We tested prototypes for material suitability, collection sufficiency, PCR compatibility, and the likelihood that they could be mass-produced, individually packaged, and sterilized quickly. Suitable prototypes were selected for an ongoing IRB-approved clinical trial. Participants were outpatients suspected of COVID-19 who presented to our drive-through test station. Each participant was swabbed with a control swab followed by a prototype swab. For each prototype, at least 10 control-swab positive and 10 control-swab negative PCR results were collected. We measured concordance using Cohen’s kappa, compared Ct values by Mann-Whitney U, and assessed staff preferences via written survey.

Results
We evaluated 45 materials and 150 designs from 23 individuals, laboratories, and companies. We have selected four so far for the clinical trial. Three have completed testing. For these, we obtained PCR results from control and prototype swabs from 230 people (74-79 pairs/prototype). Concordance was 0.85-0.89 on 10-19 control-swab positives and 58-67 control-swab negatives. Ct values were statistically indistinguishable from controls. Staff preferred two of the prototypes over the third but ultimately preferred the control swab most. The time elapsed between identification of the problem and development of a high-manufacturing-volume solution was three weeks.

Conclusions
Our experience provides lessons for how an open process can efficiently and effectively contribute to resolving a medical manufacturing crisis during a major pandemic.

Trial registration number
Beth Israel Deaconess Medical Center (BIDMC) Institutional Review Board (IRB) Protocol 2020P000323
BACKGROUND

Since the emergence of COVID-19 pandemic, more than 1.9 million cases have been reported worldwide, the result of virological testing that remains a bottleneck for controlling the pandemic. Nasopharyngeal (NP) swabs provide the highest sensitivity for detecting SARS-CoV-2 infection during early infection using commercial reverse-transcriptase polymerase chain reaction (RT-PCR)-based assays. An NP swab is an FDA Class I exempt medical device roughly 15 cm in length and 2-3 mm in diameter designed to reach the posterior nasopharynx (Fig. 1a, b). It generally consists of a shaft, a neck, and a head coated with flock (short synthetic filaments). The swab is inserted into the nasopharynx, rotated to collect material, and placed into a vial containing viral transport media (VTM). A breakpoint on the shaft enables detachment and release of the head into the vial, which is then sealed and sent for testing.

The rapid spread of SARS-CoV-2 has resulted in severe shortages of diagnostic testing supplies, including NP swabs, due to both manufacturing stoppages resulting in decreased supply and pandemic spread resulting in unprecedented demand. To address the swab shortage, hospitals and other testing centers have repurposed other commercially available swabs (e.g. throat, urogenital) to collect nasal epithelial mucosa for testing. However, these swabs are suboptimal for swabbing the nasopharynx due to differences in size and flexibility and the possibility they contain PCR-inhibitory materials.

In the short term, one solution is to design and 3D-print swabs. Advantages include simplicity over the multistep process of applying flock, the widespread availability of 3D printing capacity, and the ability to iterate prototypes rapidly. To resolve the swab-shortage crisis, we have been participating in an open collaborative process that has brought together many medical centers, individuals, academic laboratories, and both new and well established manufacturers. As part of this process, we have been testing and continuously providing feedback on prototype swabs in order to proceed rapidly but safely toward the development of swabs that can be used clinically, at volumes sufficient to address the need. The openness of the process was a conscious decision supported by a substantial body of scientific literature, including the previous experience of the present authors, that demonstrates advantages relative to closed or more siloed approaches. At our institution, this process has led to an ongoing clinical trial of several prototype swabs, the first results of which we report here.

METHODS

Preclinical evaluation (Phase I)

Design. An infectious disease physician, clinical microbiologist, and respiratory therapist tested each prototype swab for design and mechanical properties (Fig. 1c, d). This included size measurements of the head, neck, shaft, and breakpoint (requirement of ~15 cm to reach the posterior nasopharynx; head diameter of 1-4 mm to pass into the mid-inferior portion of the inferior turbinate and be able maneuver appropriately without catching in the presence of abnormal anatomy such as septal spurs or deviated nasal septum); surface properties such as smoothness vs. roughness (unpleasant feel, epistaxis risk); flexibility vs. brittleness of the head, neck, shaft, and breakpoint (to avoid fracture during use); durability (e.g. ability to tolerate 20 rough repeated insertions into a 4-mm-inner-diameter clear plastic tube curved back on itself with a curve radius of ~3 centimeters; ability to bend tip and neck 90 degrees without breaking; ability to restore to initial form following bend of 45 degrees; Fig. 1d); strength (material was required to not break with reasonable manipulation); and other factors as applicable (e.g. stickiness, smell).
Collection sufficiency. We tested for the ability to collect sufficient material for testing using a surrogate for NP swabbing and comparison to Gram stain of material obtained from Copan Diagnostics, Inc. (Mantua, Italy) model 501CS01 as the control (Fig. 1c). Slides were heat fixed and Gram stained according to the BD BBL gram stain test kit protocol. Slides were examined at 40x magnification for the presence of both epithelial cells and bacteria. Prototype swabs were passed if they collected a comparable quantity of the material as the Copan control swab.

PCR compatibility. We tested PCR compatibility by incubating the head in 3mL of modified CDC VTM (Hank’s balanced salt solution containing: 2% heat inactivated FBS, 100µg/mL gentamicin, 0.5µg/mL fungizone, and 10mg/L Phenol red) overnight to allow any PCR-inhibitory material to leach into the medium, spiking 1.5mL with 200 copies/mL (representing 2 times the limit of detection) of control SARS-CoV-2 amplicon target, vortexing, and testing using the Abbott RealTime SARS-CoV-2 Assay on an Abbott m2000 RealTime System platform, following the same protocol as for clinical testing. PCR-positive prototypes passed.

Results were recorded on a Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) spreadsheet that was kept publicly available throughout the development process on a public GitHub repository.

Production considerations (Phase II)
We considered stability to autoclaving by repeating Phase I testing on post-autoclaved materials; manufacturers’ short-term strategies for individual packaging; and manufacturers’ stated ability to produce at least 10,000 swabs per day within a week’s notice. We considered differences in supply chain to minimize the risk of future crises.

IRB and field testing (Phase III)

Trial design and oversight. COVIDSwab is an adaptive trial for evaluating the performance of prototype NP swabs (“prototypes”) as compared to Copan model 501CS01 (“control”). Participants under clinical suspicion for Covid-19 who were scheduled for standard clinical SARS-CoV-2 RT-PCR testing with a control swab were asked to be swabbed also with a single prototype. Prototypes were collected and tested until 10 positive and 10 negative results on control swabs were obtained. Sample collection was performed by trained nursing or respiratory-therapy staff (“study staff”) overseen by the respiratory therapy department at BIDMC. The Clinical Microbiology laboratories oversaw data collection. This study was reviewed and approved by the institutional review board of Beth Israel Deaconess Medical Center (protocol number 2020P000323).

Participants. Participants were individuals clinically suspected of COVID-19 who were brought to the drive-through/walk-up COVID-19 testing site at BIDMC. Adults over 18 years of age were given a participant information sheet by study staff and asked whether they would agree to being swabbed with a prototype swab performed by a trained nurse or respiratory therapist in addition to the control swab required for testing. Individuals with known thrombocytopenia of <50,000 platelets/µl were excluded from the study to avoid risk of mild bleeding.

Trial procedures. Prototype swabs were individually packaged and autoclaved for sterilization according to manufacturer protocols. Swabbing was performed per standard protocol. Participants were first swabbed with the control swab, then the prototype. Choice of naris for each swab was left to staff member and participant. Approximately half of all drive-through arrivals participated. Control and prototype swabs were placed in separate vials of VTM and
transported to the BIDMC Clinical Microbiology Laboratories where each sample was tested on the Abbott m2000 SARS-CoV-2 RT-PCR platform as per standard clinical protocol.

Statistical analyses. RT-PCR results are reported categorically as either positive or negative. We tested categorical concordance using Cohen’s kappa\(^{14}\). For each positive test, the cycle threshold (Ct) value (the RT-PCR cycle number at which the sample first turns positive) was obtained from the Clinical Microbiology Laboratories. Higher values reflect lower viral load in the sample.

We tested for systematic bias in Ct values by comparing values for controls vs. prototypes using Mann Whitney U (MWU)\(^{15}\). This tests the null hypothesis that values for controls and prototypes are drawn from the same underlying distribution. \(p>0.05\) is interpreted as no bias. For discordant (positive control/negative prototype or vice versa) samples, the negative was assigned a Ct value of 37, the total number of cycles run. As a second test for bias, we compared (again by MWU) the distribution of differences in Ct values between control and prototype swabs to the distribution of differences between two control swabs taken within 24 hours (quality-control data independent of our study). This tested the null hypothesis that the differences between control and prototype swabs and the differences between two control swabs are drawn from the same underlying distribution; \(p>0.05\) is interpreted as no bias.

To quantify relative preferences among the prototypes, we gave study staff members printouts of all six possible pairs of swabs (“round robin”), in randomized order, and for each pair asked them to circle their preference (A-B testing), collated the results, and assessed preferences.

RESULTS

Open process. A GitHub repository\(^6\) was established in the first days of the development effort to serve a public resource and knowledge base. We updated the repository continuously with design information and test results. These updates included high-resolution images of prototypes submitted to us for testing\(^6\), a public database of results of our Phase I testing, and periodic updates and guidance based on our experiences. Open communication facilitated rapid design iteration by providing anyone interested with a way to quickly understand the required specifications and to learn from each other’s experiences.

Phase I testing. To date we have evaluated 45 materials and 150 designs submitted to us for testing by 4 individuals, 2 laboratories, and 17 companies, for a total of 23 manufacturers (Fig. 1). Seven (4.7%) have passed Phase I testing. Most failures were either for inappropriate materials, including some that were sticky or brittle, or for inappropriate designs, including those with sharp heads. Prototypes from 19 manufacturers went through at least two iterations, with a maximum of 27 prototypes from one manufacturer. The rate-limiting steps were receipt of new prototypes, with slow mail delivery during the pandemic being a major contributor, and PCR-compatibility testing, as testing patient samples took priority over testing prototypes. Communication with and responsiveness by manufacturers to our feedback were considered outstanding.

Phase II and III prototypes. Four prototypes passed Phase II testing, of which three have completed our Phase III clinical trial: these are prototypes from the 3D-printing manufacturers Carbon3D, EnvisionTec, and Origin (Prototypes 1, 2, and 3, respectively; Fig. 1a). The fourth is pending clinical trial completion. Like control swabs, the prototypes were 15-16cm in length with 1-3cm length radially symmetric heads 2-3mm in diameter, a thin neck 4-7cm long and 1-2mm in diameter, and a thicker shaft 2-4mm in diameter, with a breakpoint 7-8cm from the tip of the
head. The materials were FDA-grade plastics and resins (e.g. Keysplint Soft). Head design evolved over several iterations to increase surface area. Designs generally featured a polygonal matrix connected to a central, tapered strut with multiple branch points (Fig. 1b). By varying the polygon size, manufacturers were able to balance sample collection (Fig. 1c), stiffness, and surface texture. Variations of a longitudinal central strut allowed for varying degrees of flexibility and impact cushioning (Fig. 1d).

**Sample and data acquisition.** For Prototypes 1-3, we collected and tested control and prototype swab pairs from 230 participants. Approximately half of the patients tested at the drive-through testing center participated. Because testing runs were batched and the COVID-19 status of participants was not known prior to testing, the number of control-positives sometimes exceeded the minimum requirement of 10 (range, 10-19). Total collection time was approximately three days per prototype. The frequency of positive tests was 18%.

**Comparison.** All three prototypes exhibited a high degree of concordance with the control swab, with kappas of 0.88, 0.85, and 0.89 (Prototypes 1-3, respectively; Fig. 2a). For convenience we use the terminology of true positives, true negatives, false positives, and false negatives, with the control swab result considered the provisional gold standard. Each prototype exhibited a single false positive and 1-2 false negatives. However, since control swabs are known to be an imperfect gold standard (<100% sensitivity) and because PCR positives are more likely to reflect true infection than error, false positives were interpreted as identifying missed infections. In fact per IRB protocol, false positives were referred to clinical care teams as clinically actionable. Of note, discordant cases were always associated with high Ct values, reflecting low viral burden (Fig. 2b).

To better assess possible performance differences between control and prototype swabs, we compared Ct values for control-prototype pairs for which at least one was positive (assigning the maximum-possible Ct to negatives; see Methods). Specifically, we asked whether the Ct values for the prototype swabs were systematically different from those of the control swabs. Systematically higher values for prototype swabs would suggest that they may underperform control swabs, notwithstanding the high kappa values. We found no evidence for underperformance, with MWU p-values of 0.36, 0.26, and 0.42 for Prototypes 1-3, respectively (Fig. 2b). This result indicates that the prototypes are non-inferior to the control.

As an additional assessment of non-inferiority, we compared the difference in Ct values observed between control and prototype swabs to the differences between replicates of control swabs. Independent of our clinical trial, there were 88 cases in which a patient, in the course of clinical care, was swabbed twice within 24 hours (mean±stdev, 15±7 hours), during the time period of our study. In 11 of these cases, at least one of the two swabs was positive for SARS-CoV-2. There were two disagreements between replicate swab tests (kappa=0.90), similar to what was observed in our study for each prototype (kappa=0.85-0.89). Also as in our study, the Ct values for the first swab and second swab were not significantly different (MWU p-value of 0.18). Finally, the differences between Ct values for the first and second control swabs were comparable to the differences between control and prototype swabs (MWU p-values of 0.31, 0.26, and 0.47 for Prototypes 1-3; Fig. 2b).

**Staff and participant preferences.** A written staff survey showed a preference for Prototypes 2 and 3 over Prototype 1, and a preference for the control swab over all three prototypes (Fig. 3a). In narrative feedback, Prototype 1 was described as less comfortable to participants than the others; Prototypes 2 and 3 were described as being more comparable to the control (Fig. 3b).
DISCUSSION

The COVID-19 pandemic has forced healthcare providers to seek alternative sources of critical materials affected by supply-chain disruptions and increases in demand. Over the course of our study we received numerous anecdotal reports of swab shortages at hospitals across the United States and in Europe, necessitating urgent stopgap solutions. Scientific literature on time-sensitive decentralized innovation suggests that open and collaborative processes outperform closed or proprietary ones. Here we report the success of such a process, going from the identification of the NP swab crisis to multiple clinically validated prototypes capable of high-throughput manufacture in 22 days. Notably, that none of the prototypes tested were flocked, yet their performance was indistinguishable from the flocked control.

The urgency of the situation, the configuration of the manufacturing ecosystem, and human nature contributed to several observations and shortcomings worth mentioning. First, 3D printing has important advantages in a crisis, including the ability to iterate designs and output swabs rapidly; it remains to be seen whether complementary manufacturing techniques, each with advantages and disadvantages relative to 3D printing, will prove competitive in a more mature market. Second, in any cooperative process there is a temptation to “defect,” taking without giving back; individuals and manufacturers may well exploit open knowledge for competitive advantage. This is a known price of openness that can disincentivize future cooperation, absent social or structural mechanisms to enforce norms. Third, ideally the study size would be larger, and there would be a better null model to which to compare our results, than replicates separated by many hours. Possible sources of variance in our study include differences in secretions or viral burden between nares and the possibility that the first (control) swab left little material for the second (prototype) when placed in the same naris. Despite these potential issues, our statistical tests supported analytical non-inferiority for all three prototypes. And fourth, we note the utility of our “round-robin” A-B testing survey to summarize preferences, although the narrative comments seemed often to be more positive the round robin suggested; a possible explanation is simply study staff’s familiarity with the control swab.

Like the control swab, the prototype swabs we tested can be improved upon, and manufacturers are currently doing so. The same is true for other prototypes we may test through our ongoing trial. Especially in a crisis, perfect is the enemy of good enough. We hope our experience will provide a useful case study for how to iterate and produce a clinically validated medical manufacture under the pressure of an ongoing pandemic, work on which others will hopefully improve as we continue to fight COVID-19 together. As the saying goes, it is amazing what can be accomplished when you do not care who gets the credit.

ACKNOWLEDGEMENTS

The authors would like to thank Abigail Bakken, Alisa Chau, Monique Cole, Zachary Fitts, Jonathan Goldie, Lan Huynh, Christina Lexinger-Holahan, Lorinda Longhi, Restituto Miranda, Pavel Gorelik, Jenny Hu, Ofer Mazor, Goutam Reddy, Karen Robinson, Stefan Riedel, Christina Yen, Summer Decker, Don Inger, Pawan Jolly, Kit Parker, Adama Sesay, Craig Broady, John Burpo, Daniel Davis, Joe DeSimone, Annette Friskopp, Ric Fulop, Grant Michael Gonzalez, Markus Greiner, Marie Herring, Matthew Hurley, Hardik Kabaria, David Lakatos, Ben Linville-Engler, Oren Mechanic, Richard Novak, Jifei Ou, Michael Papish, Steve Pollack, Chris Prucha, Christian Reed, Isabel Sanz, Al Siblani, Lihua Zhao, Stephanie Dominguez, Stuart Steinbock, Greg Mark, Nic Scarfo, John Bentley, Nira Pollock, Paula Watnick, Colleen Baker, Carol Daugherty. Clementina DiMonda, Virginia Dolan, Isaac Fombuh, Kylie Griffin, Amy Guadagnoli,
Allison Wang, Dan Eiref, Johan Sunryd, Jonathan Ford, Sherry Yu, and Kelsey Ladt for participation in the open process that made this study possible.

K.E.Z. was supported by a National Institute of Allergy and Infectious Diseases training grant (T32AI007061).

K.P.S. was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number F32 AI124590.

This work was conducted with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award UL 1TR002541) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Figure 1: Control and prototype swabs. (a) From left to right: the control swab (Copan 501CS01), a repurposed Hologic Aptima LBL-06850 urogenital cleaning swab approved for NP testing through our process, Prototype 1 (Carbon3D), Prototype 2 (EnvisionTec), and Prototype 3 (Origin). (b) From top to bottom, closeups of the heads of the swabs in (a). Scale bars, 1cm. (c) Examples of Gram stain of cheek swab using control (top) and prototype swabs, showing epithelial cells/cell fragments (pink) and Gram-positive bacteria (purple). Scale bar, 10µm. (d) Examples of materials testing. Clockwise from top left: head flexibility and robustness to fracture, neck flexibility and robustness to fracture, robustness to repeat insertion into and removal from a tortuous canal (diameter ~4cm), and breakpoint evaluation.
**Figure 2: Concordance results.** (a) 2x2 tables giving counts for each prototype vs. the control swab (first three panels) and for control vs. replicate control obtained within 24 hours on the same individual. Discordant results in gray; totals for each swab below and to the right of each box; total number of pairs in bold; \(K\)=Cohen’s kappa. (b) Scatterplot of Ct values for pairs of swabs for which at least one swab was SARS-CoV-2 positive. For discordant pairs, the negative swab was assigned a Ct value of 37 (the maximum number of cycles run). Note that values for each pair are not meaningfully different from each other (points cluster along the diagonal) and the pattern for each prototype vs. control data series is not obviously different from the pattern for the control vs. repeat control. These similarities are reflected by the high MWU p-values described in the Methods and Results (differences would have been reflected by low p-values).
Figure 3: Subjective feedback. (a) Round-robin A-B testing of net preferences among Prototypes 1-3 (large bold numbers) and the control ("C"). Arrows point from less preferred to more preferred swabs. Arrow weight indicates strength of relative preference. Non-bold numbers indicate the number of responses obtained for each A-B comparison. (b) Number of positive and negative comments received from study staff who administered the swabs, tabulated by comment. Negative feedback is to the left, positive feedback is to the right. Bars on both the positive and negative sides of zero reflect differing opinions among study staff. $n$, total number of comments received for each prototype. Note scales are normalized to reflect the total number of comments for ease of visual comparison.
REFERENCES


