Interactive versus static decision support tools for COVID-19: An experimental comparison

Alice Röbbelen, Malte L Schmieding, Marvin Kopka, Felix Balzer, Markus A Feufel

Submitted to: JMIR Public Health and Surveillance
on: September 23, 2021

Table of Contents

Interactive versus static decision support tools for COVID-19: An experimental comparison

Alice Röbbelen[1*] MSc; Malte L Schmieding[2*] MD; Marvin Kopka[2] MSc; Felix Balzer[2*] MSc, PhD, MD; Markus A Feufel[1*] Dipl-Ing (FH), MSc, PhD

[1]Department of Psychology and Ergonomics (IPA) Technische Universität Berlin Berlin DE
[2]Institute of Medical Informatics Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin Berlin DE
[*]these authors contributed equally

Corresponding Author:
Malte L Schmieding MD
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Charitéplatz 1
Berlin
DE

Abstract

Background: During the COVID-19 pandemic, medical laypersons with symptoms indicative of a COVID-19 infection commonly seek guidance on whether and where to seek medical care. Numerous web-based decision support tools (DSTs) have been developed, both by public and commercial stakeholders, to assist their decision-making. Though most of the DST's underlying algorithms are similar and simple decision trees, their mode of presentation differs: some DSTs present a static flowchart, while others are designed as a conversational agent, guiding the user through the decision tree's node step-by-step in an interactive manner.

Objective: To investigate whether interactive DSTs provide greater decision support than non-interactive (ie, static) flowcharts.

Methods: We developed mock interfaces for two DST (one static, one interactive), mimicking patient-facing, freely available DSTs for COVID-19 related self-assessment. Their underlying algorithm was identical and based on the Center for Disease Control's guidelines. We recruited adult US residents online. which participants. Participants appraised the appropriate social and care-seeking behavior for seven fictitious descriptions of patients (case vignettes). Participants in the experimental groups received either the static or interactive mock DST as support, while the control group appraised the case vignettes unsupported. We determined participants' accuracy, decision certainty (after deciding) and mental effort to measure quality of decision support. Participants' ratings of the DSTs' usefulness, ease of use, trust and future intention to use the tools served as measure to analyze differences in participants' perception of the tools. We used ANOVAs and t-tests to assess statistical significance.

Results: Our survey yielded 196 responses. The mean number of correct assessments was higher in the experimental groups (interactive DST group: M=11.71, SD=2.37; static DST group: M=11.45, SD=2.48) than in the control group (M=10.17, SD=2.00; $F(2,193)=8.6$, $p<.001$). Decisional certainty was significantly higher in the experimental groups (interactive DST group: M=80.7%, SD=14.1%; static DST group: M=80.5%, SD=15.8%) compared to the control group (M=65.8%, SD=20.8%; $F(2, 193)=15.7$, $p<.001$). Differences for mental effort between the three study were non-significant. Effect sizes of differences between the two experimental groups were small and non-significant for all three measures of quality of decision support and most measures of users' perception of the DSTs.

Conclusions: When the decision space is limited as is the case in common COVID-19 self-assessment DSTs, static flowcharts might prove as beneficial in enhancing decision quality as interactive tools. Given that static flowcharts reveal the underlying decision algorithm more transparently and require less effort to develop, they might prove more efficient in providing guidance to the public. Further research should validate our findings on different use cases, elaborate on the trade-off between transparency and convenience in DSTs, and investigate whether subgroups of users benefit more one type of user interface than the other.

(JMIR Preprints 23/09/2021:33733)

DOI: https://doi.org/10.2196/preprints.33733

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Röbbelen A*, Schmieding ML*, Kopka M, Balzer F*, Feufel MA*

Interactive versus static decision support tools for COVID-19: An experimental comparison

Abstract

Background: During the COVID-19 pandemic, medical laypersons with symptoms indicative of a COVID-19 infection commonly seek guidance on whether and where to seek medical care. Numerous web-based decision support tools (DSTs) have been developed, both by public and commercial stakeholders, to assist their decision-making. Though most of the DST's underlying algorithms are similar and simple decision trees, their mode of presentation differs: some DSTs present a static flowchart, while others are designed as a conversational agent, guiding the user through the decision tree's node step-by-step in an interactive manner.

Objective: To investigate whether interactive DSTs provide greater decision support than non-interactive (ie, static) flowcharts.

Methods: We developed mock interfaces for two DST (one static, one interactive), mimicking patient-facing, freely available DSTs for COVID-19 related self-assessment. Their underlying algorithm was identical and based on the Center for Disease Control's guidelines. We recruited adult US residents online in November 2020. Participants appraised the appropriate social and care-seeking behavior for seven fictitious descriptions of patients (case vignettes). Participants in the experimental groups received either the static or interactive mock DST as support, while the control group appraised the case vignettes unsupported. We determined participants' accuracy, decision certainty (after deciding) and mental effort to measure quality of decision support. Participants' ratings of the DSTs' usefulness, ease of use, trust and future intention to use the tools served as measures to analyze differences in participants' perception of the tools. We used ANOVAs and t-tests to assess statistical significance.

Results: Our survey yielded 196 responses. The mean number of correct assessments was higher in the intervention groups (interactive DST group: M=11.71, SD=2.37; static DST group: M=11.45, SD=2.48) than in the control group (M=10.17, SD=2.00). Decisional certainty was significantly higher in the experimental groups (interactive DST group: M=80.7%, SD=14.1%; static DST group: M=80.5%, SD=15.8%) compared to the control group (M=65.8%, SD=20.8%). The differences in these measures proved statistically significant in t-tests comparing each intervention group with the control group ($p<.001$ for all four t tests). The ANOVA detected no significant differences regarding mental effort between the three study groups. Differences between the two intervention groups were of small effect sizes and non-significant for all three measures of quality of decision support and most measures of users' perception of the DSTs.

Conclusions: When the decision space is limited as is the case in common COVID-19 self-assessment DSTs, static flowcharts might prove as beneficial in enhancing decision quality as interactive tools. Given that static flowcharts reveal the underlying decision algorithm more transparently and require less effort to develop, they might prove more efficient in providing guidance to the public. Further research should validate our findings on different use cases, elaborate on the trade-off between transparency and convenience in DSTs, and investigate whether subgroups of users benefit more of one type of user interface than the other.

Introduction

In December 2019, the spread of the novel lung disease COVID-19 (Coronavirus disease 2019) caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) with previously unknown etiology was detected and it developed into a global pandemic within a few weeks [1,2]. The disease courses of COVID-19 are very heterogeneous. On the one hand, it is dangerous and can be lethal even among previously healthy individuals. On the other hand, COVID-19 can present itself with very unspecific and mild symptoms [3]. Given the dynamic development of the disease and public health measures, laypersons often felt uncertain about appropriate behavior, especially when a COVID-19 infection was suspected but not proven. This situation resulted in considerable uncertainty both about what level of medical care was needed (care-seeking behavior) and whether isolation or quarantine was required (social behavior). As a result, capacity utilization in the health care system increased [4].

The increased demand for medical advice could not be met through traditional routes (eg, telephone consultation, visit to the general practitioner, visit to the emergency room, local health authorities). As a consequence, various online services with information and self-assessments (ie., patient-facing clinical decision support tools) have been developed. These decision support tools (DSTs) are comparable to commonly known symptom checkers (SC), which “are tools developed to provide support to laypersons. Users can enter their complaints and, with some SCs, demographic or health-related information (e.g., age, sex, past medical history, etc.) to obtain advice on the urgency of their complaints (triage advice) and the most likely diagnosis.” [5]. The DSTs considered in the context of COVID-19 often require information about possible exposure to the virus in addition to standard inputs related to sex, age, and major complaints. As output the DSTs provide the probability of a COVID-19 infection rather than a diagnosis. Though most of the underlying algorithms are rule-based, simple and similar decision trees, implementing official guidelines (e.g., of the Centers for Disease Control and Prevention (CDC) [6]), their mode of presentation differs: Some DSTs present the rule-based decision tree algorithm as a static flowchart [7–12], while others are designed as conversational agents (ie, similar to chatbots) with a graphical user interface, which guides the user through the decision tree’s node step-by-step in an interactive manner [13–21].

Our review of the literature suggests that there is no research, neither on SCs nor on DSTs for COVID-19, indicating whether and how the interactivity of a user interface influences the decision outcome and decision support experience. Currently available studies assessing DSTs for COVID-19 report no findings on the influence between user interaction and quality of decision support [4,22–26]. However, many publications on so-called Patient Decision Aids (PtDAs) exist that examined web-based and paper-based decision aids. In these studies, web-based PtDAs showed no difference from paper-based PtDAs in their effect on participants' decision making [27]. In terms of web-based PtDAs, interactive and static formats have not yet been compared to our knowledge.

However, the comparison between web-based tools is particularly relevant during the COVID-19 pandemic, as laypersons' decision making should ideally be supported at home to minimize the risks of unnecessary COVID-19 infections. The aim of this study is to assess whether static and interactive DSTs increase laypersons' accuracy and confidence when making COVID-19 related decisions via the Internet, and whether one type of DST (static or interactive) is superior to the other.

Methods

Study Design

We chose a between-subjects design with one three-level independent variable: Participants had to solve fictional case vignettes while receiving different types of decision support (no support in the control group, a DST in form of a static flowchart in the first intervention group or an interactive DST in the second intervention group). Participants were randomly assigned to one of the three groups. Decision support quality was evaluated by analyzing the tools' effects on decision making and the participants' judgement of the tool with multiple dependent variables: Regarding effects on decision making, the number of correct appraisals (using official CDC recommendations from 10/14/2020 as standard [28]), perceived certainty regarding the appraisal, and mental effort in making the appraisals were examined. Participants' judgments were collected on tool usefulness, ease of use, trust in the tool's recommendations, and intention to use the tool in the future.

Ethics Approval and Consent to Participate

The Ethics Committee of the Department of Psychology and Ergonomics (IPA) at Technische Universität Berlin approved this study (tracking number: ROEB_01_20200715). Participants volunteered to take part in the survey and gave informed consent before participating.

Development of the Decision Support Tools

We designed the two mock DSTs. The design was intended to mimic existing tools available on the Internet. Features of various tools related to COVID-19 were analyzed. Relevant common features of these tools were the use of short precise questions or statements to which the users could respond in agreement ("yes") or disagreement ("no"), sometimes also with uncertainty ("unsure"). These questions guided users through a decision tree's node step-by-step until a recommendation could be provided. Input requirements and output content were mostly based on recommendations from public health agencies such as the CDC. For the purpose or our study, the John Hopkins University Coronavirus Self-Checker was identified as a most suitable template because it is a decision-tree based interactive tool that requires a very small amount of information at each step as input [18]. This facilitated the construction of a second DST providing identical advice in form of a static flowchart, outlining the entire decision tree at once. In both cases the content was based on the CDC recommendations from 10/14/2020 [28].

Taking these aspects into account, we developed two mock DSTs with identical content that differed only with respect to the interactivity of their user interface: A static flowchart where users could follow the arrows of the appropriate path using their gaze (Supplemental 1). This flowchart was developed using Microsoft PowerPoint [29]. The other tool was implemented as a simple interactive conversational agent, where users clicked on buttons to respond to questions, guiding the user through the nodes of the decision tree step by step (Supplemental Figure 2). The individual screens of the interactive DST were also designed using Microsoft PowerPoint and then linked together using InVision [30], enabling dynamic interaction. Mimicking the output of existing tools, five different recommendations were possible: "emergency care is required", "selfcare and physical distancing are sufficient", "selfcare is sufficient and quarantine necessary", "selfcare is sufficient and isolation is necessary" and "non-emergency care and isolation are required". We anticipated that laypersons would struggle comprehending the difference between "quarantine" and "isolation", yet decided to differentiate between them, as some COVID-DST used these terms without explaining them [7,11]. During data analyses, however, we rated answers as correct when participants confused these two terms (see Data Analysis section). We developed mock tools rather than choosing existing DSTs from the Internet to ensure comparability of algorithm and design of the tools, as well as stable integration into the survey throughout the surveying period.

Preparing the Case Vignettes

We created seven short patient descriptions (case vignettes). In this process, we extracted decision criteria identified by the CDC and their recommendations concerning help-seeking and social behavior in case of a potential COVID-19 infection [28]. Decision criteria included the presence of typical COVID-19 symptoms with the expressions: (0) no symptoms, (1) primary symptoms (typical for a COVID-19 infection), and (2) secondary symptoms (can also occur in COVID-19 infection but are less typical) [28] (see Supplemental Table 1). Another decision criterion represents the potential contact with the pathogen SARS-CoV-2 and the expressions: (1) critical contact with a confirmed infected person has occurred, (2) contact may have occurred, or (3) critical contact can be excluded. The last important criterion is the presence or absence of risk factors classified by the Robert-Koch-Institute (RKI), Germany's federal public health authority [3] (eg, certain pre-existing conditions such as chronic lung disease; see Supplemental Table 2). Through a pre-test with 15 participants, in which we evaluated the response behavior of the participants as well as their free text feedback, we revised the wording of the vignettes improving intelligibility and removing errors spotted by the pre-test participants. The vignettes included more information than necessary for the appraisal to better simulate a real-life decisional context and increase ecological validity. We also ensured that all information asked for in the developed DSTs was included in the vignettes, and that the fictitious patients represented diverse age and gender groups. In addition, each possible outcome of the mock tools was covered at least once as a correct solution. The appendix contains an outline of the seven case vignettes (see Supplemental Tables 3 and 4).

Data Collection

Participants

All participants lived in the United States (US), were at least 18 years old, and had no professional medical background. Our investigation was limited to US residents, as recommendations of the developed DSTs were based on CDC recommendations and guidelines and thus applied to the US region.

Survey and Instruments

We created an online survey with UNIPARK (QuestBack GmbH) [31]. To ensure differences between the three groups are not due to confounding variables we surveyed variables we suspected to have an influence on the main outcome measures, namely, sociodemographic factors, affinity towards interacting with technology, prior knowledge on COVID-19, prior use of a COVID-19 DST and perceived threat of a COVID-19 infection. The sociodemographic variables consisted of age, sex, US residency, and highest level of completed formal education. Affinity towards interacting with technology was surveyed with the short version of the Affinity for Technology Interaction Scale (ATI-S) [32]. The perceived threat of COVID-19 was surveyed with an instrument by Kim and Park [33]. Participants' prior knowledge on COVID-19 necessary to appraise the adequate health-seeking and social behavior for a suspected COVID-19 patient was assessed with five self-developed, multiple-choice questions (see Supplemental Figure 3).

In the survey's main part participants were presented the seven case vignettes in random order. In both the experimental groups and the control group, participants provided their personal appraisal of the adequate help-seeking behavior for each case vignette and, in a further question, of the adequate social behavior. In the two experimental groups participants were also prompted to state which recommendations of the tool (supposedly) provided concerning help-seeking behavior and on social behavior (see Supplemental Figures 4 and 5).

As we anticipated participants commonly erring in determining the appropriate social behavior when required to differentiate between "isolation" and "quarantine", we deemed both answers correct for our main analysis and conducted a second analysis without this adjustment (Supplemental Figure 9).

After each case vignette, participants were further asked to rate their mental effort required in making decisions concerning the respective fictitious case presentation using a 9-point category scale ranging from "very, very low mental effort" to "very, very high mental effort" (see Supplemental Figure 6) [34]. This scale by Paas and colleagues can be considered a "subjective, indirect measure of cognitive load" [35] in making decisions regarding case vignettes. Since mental effort increases with higher perceived demands of a stimulus or task [35], it was included here as an indicator of the quality of decision support. We assume that a good DST guides people to the right decision without requiring high cognitive load.

Following all seven case vignettes, participants' decisional uncertainty was assessed once using O'Connor's (1995) Decisional Conflict Subscale [36]. On three items with a five-point scale, participants rated how confident they were in their decisions, see Supplemental Figure 7. The scale was developed to evaluate health-care consumer decision aids. We used this scale as an indicator of the quality of decision support, since we assume that a good DST helps medical laypersons to feel confident in their decisions. This assumption is supported by O'Connor's claim that "decision aids should reduce uncertainty and confusion in choosing a course of action" [36].

In addition to the effect of the tools on decision making and participants' perceptions of their decision making, we also asked directly about participant's perceptions of the tools in the two intervention groups. They were asked about the perceived usefulness and ease of use of the tool after having worked on all seven cases. Perceived usefulness is defined as the individual's perception of the extent to which using the tool improves decision-making performance. Perceived ease of use implies that using the tool does not require any effort [37]. Both constructs were measured using scales from Davis' Technology Acceptance Model (1989) [38]. In addition, the intentions to use the DST in the future and the trust in the recommendations of the DST were measured as a subjective rating using one item each (Supplemental Figure 8 presents the exact phrasing of the items).

Procedure

We recruited participants via the crowdsourcing platform Prolific.co [39]. This platform is aimed at researchers and allows them to conveniently recruit participants for online surveys and experiments. Prolific.co provides a pool of participants that can be screened based on demographic data. Researchers can make their survey or experiment available on Prolific.co and users registered on Prolific.co as prospective participants can choose whether to participate in studies which they are eligible for according to the set pre-screening criteria. We chose Prolific.co as it is characterized by a diverse population that provides high quality data [40]. Each participant received a remuneration of US $2.00 for completing the survey. In addition, participants could earn a bonus of $0.20 for each correct decision. To ensure that all participants answered the questions attentively, attention checks were added to the survey. Participants failing more than one of three attention checks were excluded. Furthermore, at the end of the survey, participants were asked to self-report whether they had participated in the survey attentively, honestly, and without external assistance as suggested by Rouse [41]. Participants were remunerated independently of their answer, but only data from participants confirming this question were included in the analysis. By selecting the weekend day and early afternoon PDTs, we attempted to recruit a population as diverse as possible as suggested by Casey et al. [42]. The survey was released for participation on Prolific.co on four different days (November 21, 2020, at 1 PM Pacific Daylight Time (PDT); November 22, 2020, at 11 AM PDT; November 28, 2020, at 12 PM PDT; and November 29, 2020, at 1 PM PDT). On each day 50 participants were recruited within a few hours after release.

Data Analysis

Data were cleaned and explored using R 4.0.2 [43] and the tidyverse packages [44]. After examining the central tendencies and distributions of the variables separately for the three groups and plotting them using the package ggplot2 [45], we performed a one-way analysis of variance to compare the three conditions, separately for the dependent measures accuracy, decisional certainty, and mental effort. If significant group differences were present, we followed up with Bonferroni-corrected post-hoc pairwise t-tests. For the dependent variables usefulness, ease of use, trust in tool recommendations, and future use intention, we conducted Two Sample Welch t-tests to compare the two intervention groups because the sample sizes of the groups were unequal [46]. Effect sizes were quantified with Cohen's *d* and calculated using the rstatix package [47]. When results were not statistically significant, we performed a post-hoc power analysis using the pwr package [48] with the corresponding group sizes, an alpha level of α = .05 and a power of 1-β = .80.

Results

Participant Characteristics

In total, our survey was accessed 233 times during the 4 days it was available. 37 participations could not be used for data analysis because participants either dropped out of the questionnaire (n=12), exceeded the maximum survey completion time (n=10), failed attention checks (n=6), or did not meet eligibility criteria (n=9). The remaining participants all affirmed that they had paid close attention during the survey and answered honestly. This yielded a total of 196 participants, who have assessed all 7 case vignettes and could be included in the analysis (see Table 1 for details). The mean time for completion of the survey was 22 minutes and 3 seconds. Across the three groups the participant characteristics were very similar, see Table 1.

Characteristics	**Values of the total sample**	**Group 1: "Control Group (No DST)"**	**Group 2: "Static DST"**	**Group 3: "Interactive DST"**
Sample size: n	196	66	62	68
Age in years: median (IQR)	30 (18)	30 (17.2)	26.5 (13.2)	33 (20.5)
Gender: n (%)				
- Female	94 (48)	31 (47)	27 (44)	36 (53)
- Male	100 (51)	33 (50)	35 (56)	32 (47)
- Other	2 (1)	2 (3)	0 (0)	0 (0)
Education: n (%)				
- Non–high school graduate	4 (2)	1 (2)	1 (2)	2 (3)

- High school graduate	34 (17)	9 (14)	18 (29)	7 (10)
- Some college	66 (34)	22 (33)	20 (32)	24 (35)
- Bachelor's degree	49 (25)	20 (30)	12 (19)	17 (25)
- Graduate degree or higher	43 (22)	14 (21)	11 (18)	18 (26)
Recent experience with COVID-19 decisions [a]**:** n (%)				
Recently faced triage decision	70 (35)	25 (38)	27 (44)	18 (26)
Recently faced social behavior decision	101 (52)	35 (53)	32 (52)	34 (50)
Recently consulted a static DST to face the decision	56 (29)	15 (23)	21 (34)	20 (29)
Recently consulted an interactive DST to face the decision	53 (27)	21 (32)	15 (24)	17 (25)
Medical training: n (%)				
- No training	163 (83)	55 (83)	51 (82)	57 (84)
- Basic first aid training	33 (17)	11 (17)	11 (18)	11 (16)
Affinity for technology interaction on a scale of 1 to 6: median (IQR)[b]	4 (1.2)	4 (1.7)	4 (1)	4 (1)
Perceived threat of COVID-19 on a scale of 1 to 7: median (IQR)[c]	5 (2)	5.25 (1.6)	5 (2.1)	5 (2)
Prior knowledge of COVID-19 on a scale of 0 to 5: median (IQR)[d]	3 (2)	3 (2)	3 (1)	3 (1)

Table 1. Participant characteristics (N=196) of an experimental study assessing the influence of decision support tools (DSTs) s on laypersons' COVID-19 related appraisals. Participants were non-medically trained US inhabitants sampled online in November 2020. [a] Recent was defined as "in the last 6 months". [b] Measured by Wessel's Affinity for Technology Interaction Short Scale (ATI-S). [c] Measured by a subjective self-assessment on two items on a scale of 1 to 7 adapted from Kim and Park. [d] Measured by the number of correctly answered multiple choice questions with reference to COVID-19.

Effects on decision making

The omnibus ANOVA detected group differences with respect to decision accuracy (overall and separately for decisions about help-seeking behavior and social behavior) and perceived certainty, but not for mental effort in decision making, see Table 2.

Dependent variables	Group 1: "Without DST"	Group 2: "Static DST"	Group 3: "Interactive DST"	Test statistics of group comparison
Accuracy: mean (sd)				
Total number of correct decisions (min = 0; max = 14) [a]	10.17 (2.00)	11.45 (2.48)	11.71 (2.37)	F (2, 193) = 8.59, P < .001, η^2=.08
- Number of correct decisions on help-seeking behavior	4.82 (0.96)	5.47 (1.39)	5.54 (1.40)	F (2, 193) = 6.58, P < .001, η^2 = .002

(min = 0; max = 7)				
- Number of correct decisions on social behavior (min = 0; max = 7) [a]	5.35 (1.41)	5.98 (1.26)	6.16 (1.18)	F (2, 193) = 7.33, P < .001, η^2 = .02
Decisional certainty Certainty Score from 0 to 100 [b]: mean (sd)	65.78 (20.78)	80.51 (15.89)	80.72 (14.08)	F (2, 193) = 15.67, P < .001, η^2=.14
Mental Effort on a scale from 1 to 9**:** mean (sd)	5.62 (1.57)	5.40 (1.68)	5.09 (1.78)	F (2, 193) = 1.73, P = .18

Table 2. Omnibus ANOVAs measuring the effects of different decision support tools (DSTs) on laypersons' ability to correctly appraise fictitious descriptions of patients with symptoms indicative of COVID-19 in an experimental study. 196 study participants (all US residents and non-medically trained) were recruited online in November 2020 and asked to judge how fictitious patients with symptoms indicative of COVID-19 should behave. Participants were randomly assigned to one of three groups in which they received either support by a static DST (flowchart), an interactive DST (mimicking a conversational agent) or no support. . [a] The response options "quarantine" and "isolation" were not differentiated for this analysis, that is, they were both considered correct because layperson participants commonly confuse these terms. An analysis without this adjustment is provided in Supplemental Figure 9 and shows the same trend. [b] Responses were transformed into a Certainty Score between 0 and 100., i.e., 0 = "person feels extremely uncertain about best choice" and 100 = "person feels extremely certain about best choice".

Accuracy

Across all groups participants' appraisals were commonly correct. The average participant decided correctly in more than 70% (10/14) of the decisions. Overall decision accuracy was higher in the experimental groups receiving support from a DST than in the unsupported control group, see Figure 1 and Table 2. Bonferroni corrected pairwise Welch Two Sample t-tests indicated statistically significant differences with moderate effect sizes between control and experimental groups (Control group v Static DST: t(117.35) = -3.21, P <.001, Cohen's d = -0.57 [CI: -1.1 to –0.14]; Control Group v Interactive DST (t(129.6) = -4.06, P <.001, Cohen's d = -0.70 [CI: -1.31 to –0.21]). The difference in decision accuracy between the two experimental groups showed a negligible effect size and was not significant (t(125.58) = -0.59, P = .55, Cohen's d = -0.10 [CI: -0.54 to 0.34]). Based on a post-hoc power analysis we estimate that, on a population level, the real difference in accuracy is less than Cohen's d = 0.5 (equivalent to 1.22 correct responses) with a power of 0.8.

We found similar results when analyzing participants' accuracy for help-seeking behavior and social behavior separately, see Table 2. When the differences between the response options "isolation" and "quarantine" are considered, the pattern remains the same, but the gap in accuracy between experimental and control groups widens, see Multimedia Appendix 13.

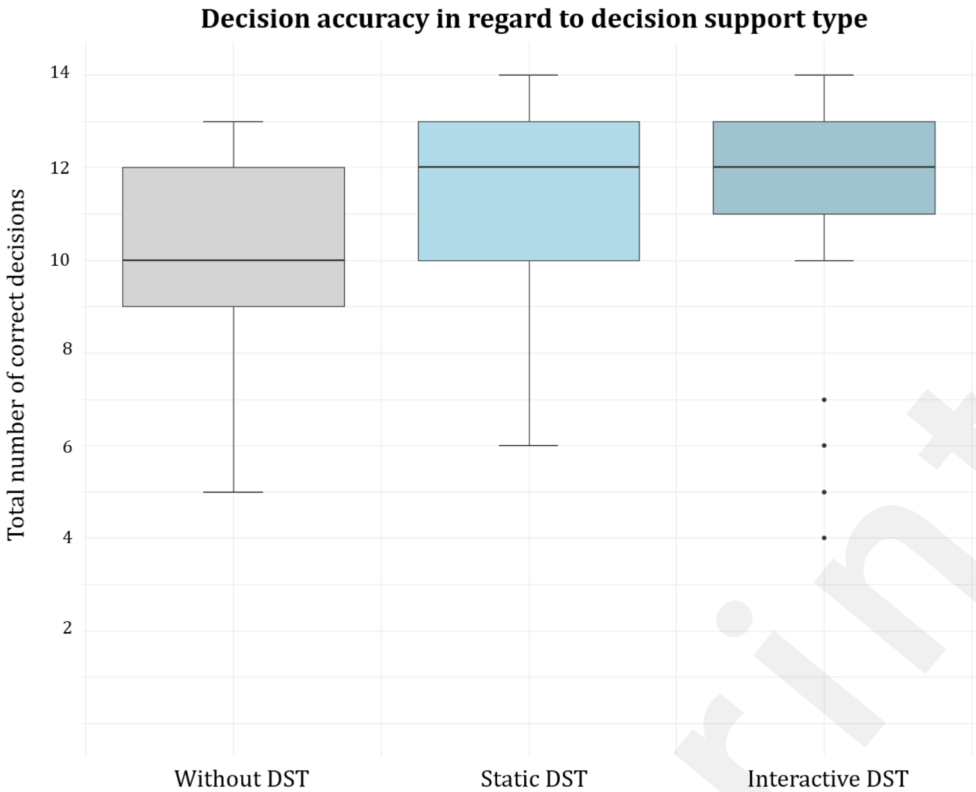


Figure 1. Boxplot showing the distribution of the 196 participants' decision accuracy to appraise 7 fictitious descriptions of patients with symptoms indicative of COVID-19. Study participants (all US inhabitants, non-medically trained, sampled online in November 2020) were tasked to answer two questions per patient description. We randomly assigned participants to one of three experimental groups; in two groups they were supported by either a static decision support tool (DST), ie a flowchart, or an interactive DST, ie a conversational agent mimicking a chatbot. In the control group, they received no decision support. . The boxplots' filled box represents the interquartile range (IQR), the horizontal line inside the box the median, the whiskers the maximum and minimum values within 1.5 IQR of the median, and single dots represent outliers of participants' total number of correct decisions.

Decisional certainty

Participants were commonly certain in their decision making. Less than 8% (15/196) indicated a certainty of less than 50 on a scale from 0 to 100, in contrast to 16% (32/196) indicating maximum certainty (100/100). Participants' certainty in their decisions differed between the three groups: Certainty in decision making was rated lower by participants without decision support than by those receiving decision support, see Figure 2 and Table 2. Bonferroni corrected pairwise Welch Two Sample t-tests mark these differences as statistically significant with large effect sizes (Control group v Static DST: t(121.1) = -4.51, *P* <.001, Cohen's *d* = -0.79 [CI: -1.22 to -0.37]; Control Group v Interactive DST (t(115.67) = -4.67, *P* < .001, Cohen's *d* = -0.81 [CI: -1.27 to –0.44]).

Decision certainty in the two experimental groups was nearly identical with mean certainty scores of 80.51 (Static DST) and 80.72 (Interactive DST). The inferential analysis indicates this difference to be of negligible effect size non-significant (t(123.68) = -0.09, *P* = .92, Cohen's *d* = 0.01 [CI: -0.47 to 0.43]). Based on a post-hoc power analysis we estimate that, on a population level, the real difference in decisional certainty between both experimental groups is less than Cohen's $d = 0.5$ (equivalent to 7.5 percentage points) with a power of 0.8.

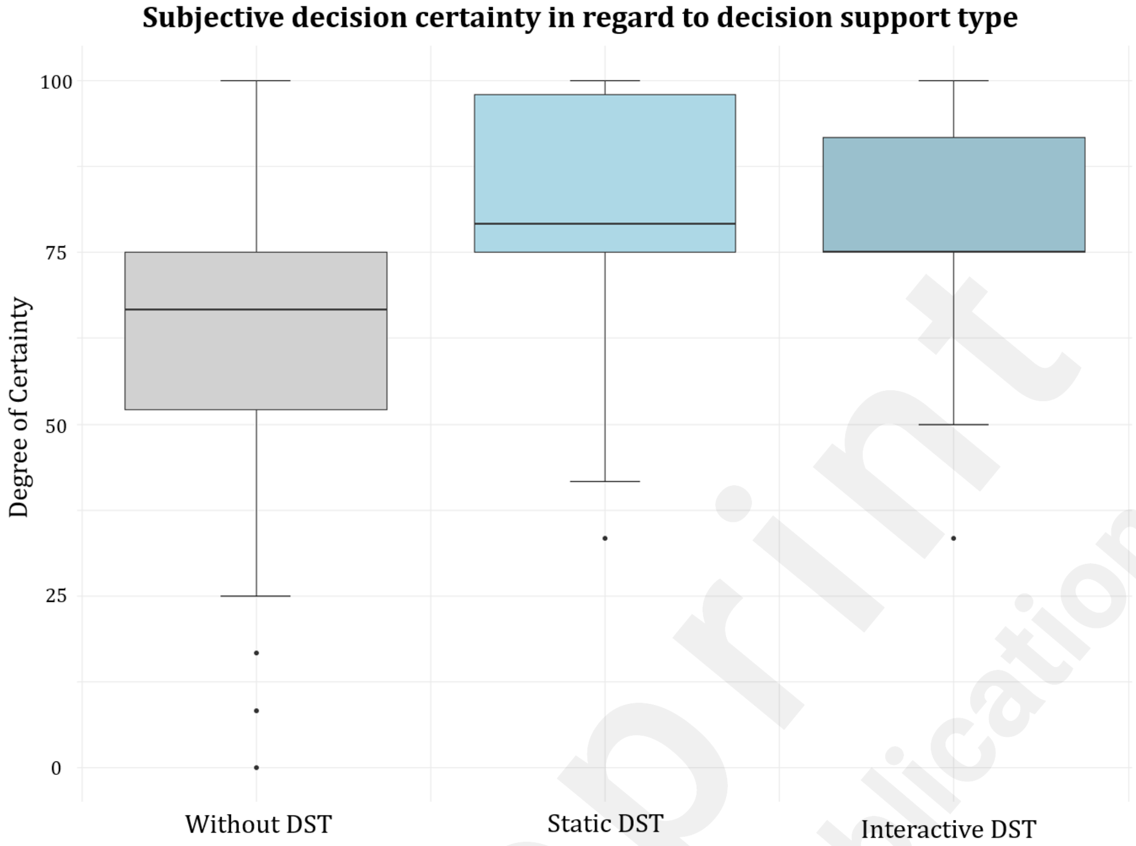


Figure 2.
Laypersons' perceived certainty in their own appraisals of COVID-19 related clinical decisions obtained in an experimental study in November 2020. The 196 study participants were US residents, non-medically trained and sampled online. Our study tasked them to assess 7 fictitious descriptions of patients with symptoms indicative of COVID-19. Participants were randomized to either receive support from a static decision support tool (DST), ie a flowchart, or an interactive DST, ie a conversational agent mimicking a chatbot. Following the 14 appraisals, we surveyed the participants' certainty in their answers using the Decisional Conflict Scale. A score of 0% indicates minimum certainty, while 100% indicates maximum certainty. The boxplots' filled box represents the interquartile range (IQR), the horizontal line inside the box the median, the whiskers the maximum and minimum values within 1.5 IQR of the median, and single dots represent outliers of participants' total number of correct decisions.

Mental effort

The perception of how demanding the task was varied among participants. While a third (63/196) stated that the tasks required low mental effort (ie, they indicated an average mental effort score below 4.5 on a scale from 1 to 9), four out of ten considered to task to be at least somewhat demanding (indicating scores above 5.5).

Participants rated the mental effort of decision making lowest in the Interactive DST group (mean = 5.06) and highest in the control group unsupported by a DST (mean = 5.62), see Table 2. However, the differences between group means were not significantly in the omnibus ANOVA, see Table 2. Based on a post-hoc power analysis we estimate that, on a population level, the effect size of the difference is not greater than $\eta^2 = 0.047$ with a power of 0.8.

Participants' perceptions of the tools

We investigated participants' perceptions of the mock tools with four metrics: perceived usefulness, perceived ease of use, trust, and intention to use the tool again in the future. On average, participants perceived the DSTs positively: their self-reported trust in the tools was generally high in both experimental groups as was the number of DST recommendations participants followed (on average in more than 12 out of a total of 14 recommendations), see Table 3. Accordingly, almost all (95%, 126/130) participants judged the mock tools to be useful (ie, indicated a Perceived Usefulness score of above 3.0 on a scale from 1 to 5) with half of the participants scoring the usefulness at or above 4.75. Ease of use was rated similarly positive (see Table 3) and their intention to use the tool in the future was high on average. Regarding differences between the two DST, the mock users' scores on perceived usefulness, trust and future use intention were higher in the static than the interactive DST group. Ease of use was perceived higher in the interactive DST group. However, all effect sizes of differences between the two DST groups were small (Cohen's d <= 0.35) and proved significant only for Future use intention in a Two sample Welch t-test, see Table 3.

Based on a post-hoc power analysis we estimate that, on a population level, the real difference in these variables between both groups is less than Cohen's d = 0.5 (equivalent to a difference of 0.34 for perceived usefulness and perceived ease of use, a difference of 0.42 for self-reported trust, 0.96 more followed decisions and a difference of 0.51 for future use intention) with a power of 0.8.

Dependent variables	Group 2: "Static DST"	Group 3: "Interactive DST"	Test statistics of group comparison
Perceived Usefulness on a scale from 1 to 5: mean (sd)	4.56 (0.65)	4.35 (0.71)	t (128) = 1.7, P = .091, d=0.30
Perceived Ease of Use on a scale from 1 to 5: mean (sd)	4.26 (0.71)	4.47 (0.58)	t (117.92) = -1.90, P = .059, d=-0.34
Trust: mean (sd)			
Self-reported trust in the tool's recommendation on a scale from 1 to 7	6.08 (0.84)	5.83 (0.85)	t (127.15) = 1.74, P = .083, d = 0.31
Total number of decisions where the recommendation of the tool was followed on a scale from 0 to 14	12.73 (1.81)	12.5 (2.03)	t (127.94) = 0.67, P = .748, d = 0.12
Future use intention on a scale from 1 to 7: mean (sd)	6.23 (0.88)	5.87 (1.16)	t (123.91) = 1.99, P = .048, d=0.35

Table 3. Laypersons' perceptions of two mock decision support tools (DSTs) for COVID-19 related clinical decisions obtained in an experimental study in November 2020. The 196 study participants were US residents, non-medically trained and sampled online. Our study tasked them to assess 7 fictitious descriptions of patients with symptoms indicative of COVID-19. Participants were randomized to either receive support from a static DST, ie a flowchart, or an interactive DST, ie a conversational agent mimicking a chatbot, or no support (control group).

Subsequently, the participants in the interventional groups were asked to rate the given tool's usefulness and perceived ease of use, state their trust in the tool and their future intention to use the tool. We measured usefulness and perceived ease of use according to Davis' Technology Acceptance Model.

Discussion

Principal Findings

In our study, DSTs increased laypersons' accuracy and certainty in decision-making. Potentially, DSTs can also reduce the mental effort in decision making, but this effect was statistically non-significant in our experiment. Thus, our experiment confirms the benefit of DSTs for COVID-19 related self-triage discussed in prior studies [4,23].

Regarding the question whether one mode of presentation of DSTs (ie, static or interactive presentation) is superior to the other, our experiment produces evidence that differences in measures of quality of decision support and participants' perception of the tools are small. Interactive DSTs are potentially more convenient to use. Users of the interactive DST rated the mental effort as lower and the perceived ease of use as higher than the users of the static DST. However, these differences were not significant and did not translate into a higher level of trust, greater perceived usefulness or greater intention to use the tool in the future. On the contrary, users rated the static tool more favorably on these measures than the interactive DST. Overall, the effect sizes for differences in these measures were low and statistically non-significant.

To our knowledge our study is among the first to directly compare the effectiveness of different modes of presentation on the quality of support received from web-based patient-facing clinical decision support tools. Our results are in line with findings from similar research in a different field of application: In their 2018 meta-analysis on the use of PtDAs to support decisions concerning prostate cancer screening, Baptista et al. [27] concluded that web-based decision aids reduce decisional conflict compared to no decision aid, but no more than static, print-out decision aids. Comparability between Baptista et al. [27] and our study is limited, as web-based decision aids and their paper-based printed-out versions assessed by Baptista et al. [27] do not directly match the mock conversational agent and the static flowchart we presented digitally in our study. However, both results suggest that, for a decision scenario with clear-cut options, the mode of presentation with a decision support tool has little to no effect on how helpful it is for their users.

Practical Implications

First, our results underline the benefits of making DSTs available to laypersons for decisions with clear-cut options such as those encountered in the COVID-19 pandemic, as both decisional accuracy and trust in the decision increased when laypersons were supported by both mock

DSTs. Second, the quality of decision support provided by the static flowchart and the interactive tool did not differ significantly and post-hoc sensitivity analyses on effect sizes indicate that we can rule out large effects.Thus, factors like development effort might potentially weigh more than format of interaction when deciding on how to present decision support to laypersons

When complexity is low, as in the case of the COVID-19 decisions, the static version may provide a better overview and thus make the decision more transparent to its users and be quicker and more cost-efficient to develop and publish than an interactive conversational agent. On the other hand, for decisions with higher complexity, full transparency might compromise ease of use of a static DST. Thus, for more complex decisions, interactive tools that guide the user step by step may be more user friendly. However, our results suggest that interactivity is not an effective means in itself to increase the usefulness of a DST: First, the higher convenience of using the interactive DST did not translate into greater trust or perceived usefulness, nor did it yield a greater increase in decision accuracy or certainty than static flowcharts. As the latter entail less development efforts while increasing transparency, they might be the preferable mode of DST for public health officials to implement on their websites to provide guidance to the public on decisions of low complexity and limited decision space.

Our study also raises two topics for further research: First, which factors increase users' trust in DSTs and how they are weighed. From our results we can only speculate that users prefer transparency over convenience. Second, whether subgroups of the population might prefer one type of interface over the other. While our results showed only minor differences between sample averages of users of the static and interactive DSTs, potentially subgroups of the population benefit more from one type of interface than the other. For example, users with low technological affinity might prefer a static flowchart over a conversational agent.

Limitations

The results from our study are mainly limited by concerns about external validity. That is, the interactive mock DST we developed does not fully represent the whole variety of interactive tools that are available on the Internet. Although our DST mimics the interactive tool from Johns Hopkins University, other tools incorporate significantly more decision factors and possible outcomes. Secondly, the sampled study population's composition of educational background is not representative of the adult US population. Our study sample included a higher proportion of highly educated persons, with only 2% (4/196) being non-College graduates as compared to 55% in the US adult population [49]. Also, the mock interactive tool we developed for this article does not fully exhaust the potential of interactivity as our participants only followed prompts to respond to binary questions by clicking either “Yes” or “No” buttons, while more interactive tools also require users to enter information manually (eg, age or current location). Finally, participants in this study did not use the tool for self-assessment, but to appraise cases with fictitious patients. This means that they did not experience personal concern, as would likely be the case with COVID-19 suspicion in a real-world use situation. To promote external validity, the recommendations of the DSTs we developed conform with all CDC guidelines and its interactive capacities are basic but mimic those of existing DSTs.

Conclusions

When the decision space is limited, a static flowchart potentially performs just as well as an interactive tool in enhancing the decision quality of laypersons with symptoms indicative of a COVID-19 infection. As static flowcharts reveal their underlying decision algorithm more transparently, they might prove to be more suitable in not only guiding laypersons through the healthcare system, but also in communicating the reasoning and thereby empowering patients. Further research should validate our findings on different use cases, elaborate on the trade-off between transparency and convenience in DSTs, and investigate whether subgroups of users benefit more one type of user interface than the other by assessing interactions between outcome variables (eg, accuracy, mental effort, perceived usefulness etc.) and participant characteristics (eg, age, eHealth literacy). We have made all data necessary to conduct these exploratory analyses and to reproduce our reported findings publicly available [50].

Acknowledgments

The authors express their gratitude to the participants and to Eike Richter for his advice on statistical methods. The project was funded by the home institutions of the last authors (MF and FB). No external funding was required to conduct this study. The authors acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité - Universitätsmedizin Berlin.

Authors’ Contributions

AR and MLS conceived the study. AR created the questionnaire, developed the mock DSTs, designed and conducted the analyses, and wrote the first draft of the paper. MLS oversaw the creation of the case vignette adaptations and aided in the design of the methods and manuscript development. MK assisted in statistical analyses and manuscript development. MF and FB provided critical input and advised on the study and questionnaire design, analysis methods, and drafts of the paper. AR and MLS share the first authorship. MF and FB share the last authorship. All authors accept full responsibility for the final version of the paper.

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Conflicts of Interest

All authors have completed the International Committee of Medical Journal Editors uniform disclosure form and declare no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Abbreviations

CDC: Center of Disease Control
DST: Decision Support Tool
SC: Symptom Checker
WHO: World Health Organization

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery 2020 Apr;76:71–76. [doi: 10.1016/j.ijsu.2020.02.034]
2. World Health Organization. Pandemie der Coronavirus-Krankheit (COVID-19) [Internet]. 2020 [cited 2020 Jun 24]. Available from: https://www.euro.who.int/de/health-topics/health-emergencies/coronavirus-covid-19
3. Robert Koch Institut. Epidemiologischer Steckbrief zu SARS-CoV-2 und COVID-19: 8. Demografische Faktoren, Symptome und Krankheitsverlauf [Internet]. 2021 [cited 2021 May 2]. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html
4. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, Moriarty T, Gleason N, Intinarelli G, Gonzales R. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. Journal of the American Medical Informatics Association 2020 Jun 1;27(6):860–866. [doi: 10.1093/jamia/ocaa051]

5. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study. J Med Internet Res 2021 Mar 10;23(3):e24475. [doi: 10.2196/24475]
6. Centers for Disease Control and Prevention. What to Do If You Are Sick [Internet]. 2021 [cited 2021 May 5]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/steps-when-sick.html
7. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) Risk Assessment and Public Health Management Decision Making [Internet]. 2020 [cited 2021 May 2]. Available from: https://dhr.idaho.gov/wp-content/uploads/2020/COVID-19/CDCDecisionMakingModel.pdf
8. Jackson Free Press. COVID-19 Flowchart [Internet]. 2020 [cited 2020 May 2]. Available from: https://www.jacksonfreepress.com/photos/2020/mar/14/36329/
9. North Dakota Department of Health. What To Do If You Are A Close Contact [Internet]. 2020 [cited 2020 Oct 9]. Available from: https://www.health.nd.gov/what-do-if-you-are-close-contact
10. Robert Koch Institut. COVID-19: Bin ich betroffen und was ist zu tun?: Orientierungshilfe für Bürgerinnen und Bürger [Internet]. 2020 [cited 2021 May 2]. Available from: https://edoc.rki.de/bitstream/handle/176904/6779.2/Orientierungshilfe_COVID_Buerger_20-05-06_doi_10.25646-6629-2_korr.pdf?sequence=1&isAllowed=y
11. Michigan.gov. Contact Tracing Flow Chart: For potential and actual exposures to COVID-19 [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.michigan.gov/documents/coronavirus/Person_Exposure_Final_3-25-2020_684832_7.pdf
12. Coronavirus-Entscheidungshilfe [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.zeit.de/wissen/gesundheit/2020-03/Coronavirus-Entscheidungshilfe-2020-03-31.pdf.
13. Ada Health GmbH. COVID-19 Screener [Internet]. 2020 [cited 2021 May 2]. Available from: https://ada.com/de/covid-19-screener/#assessment
14. Centers for Disease Control and Prevention. Coronavirus Self-Checker [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html
15. Charité Universitätsmedizin Berlin. CovApp: Haben Sie Symptome? CovApp zeigt Ihnen mögliche nächste Schritte [Internet]. 2020 [cited 2021 May 2]. Available from: https://covapp.charite.de/
16. Emory University School of Medicine. Coronavirus Checker [Internet]. 2020 [cited 2021 May 2]. Available from: https://c19check.com/start
17. in4medicine AG. COVID-GUIDE [Internet]. 2020 [cited 2021 May 2]. Available from: https://covidguide.health/de/
18. The Johns Hopkins University; The Johns Hopkins Hospital; The Johns Hopkins Health System Corporation. Coronavirus Self-Checker and COVID-19 Vaccine FAQ [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.hopkinsmedicine.org/coronavirus/covid-19-self-checker.html
19. Universitätsklinikum Giessen und Marburg GmbH. COVID-ONLINE [Internet]. 2020 [cited 2021 May 2]. Available from: https://covid-online.de/#/
20. USC Gehr Family Center for Health Systems Sciences & Innovation. Covid assessment [Internet]. 2020 [cited 2021 May 2]. Available from: https://covidassessment.org/acknowledgement
21. WebMD. COVID-19 Symptom Checker [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.webmd.com/coronavirus/coronavirus-assessment/default.htm
22. Kahnbach L, Lehr D, Brandenburger J, Mallwitz T, Jent S, Hannibal S, Funk B, Janneck M. Quality and Adoption of COVID-19 Tracing Apps and Recommendations for Development:

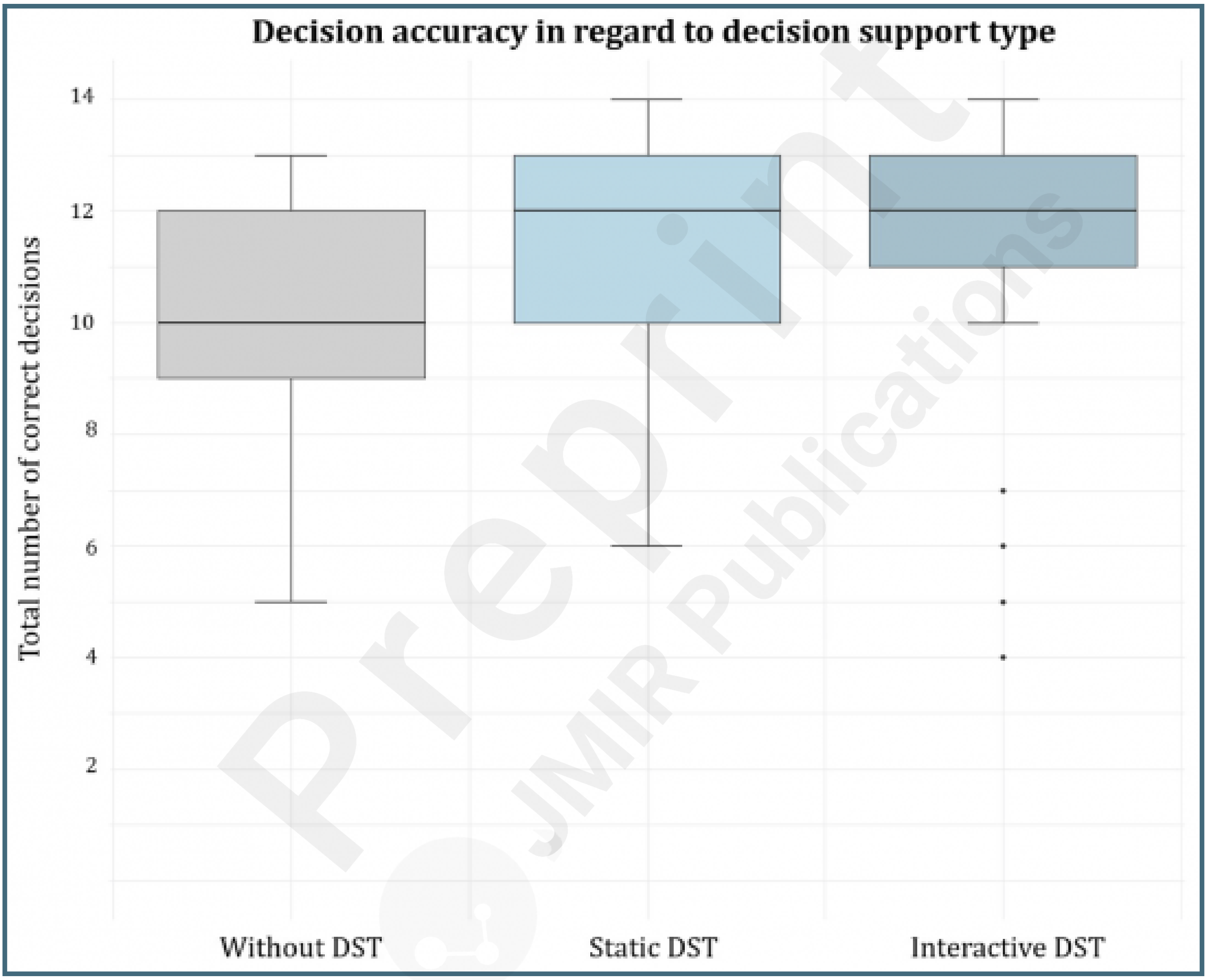
Systematic Interdisciplinary Review of European Apps. J Med Internet Res 2021 Jun 2;23(6):e27989. [doi: 10.2196/27989]
23. Lunn PD, Timmons S, Julienne H, Belton CA, Barjaková M, Lavin C, McGowan FP. Using decision aids to support self-isolation during the COVID-19 pandemic. Psychology & Health 2021 Feb 1;36(2):195–213. [doi: 10.1080/08870446.2020.1849701]
24. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, Knapp B. Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study. J Med Internet Res 2020 Oct 6;22(10):e21299. [doi: 10.2196/21299]
25. Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, Jeffries HE, Fisher E, Steuerman Y, Namir Y, Goldschmidt Y, Souroujon D. Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study. J Med Internet Res 2020 Oct 20;22(10):e23197. [doi: 10.2196/23197]
26. Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-Based Tracking of Self-Reported COVID-19 Symptoms: Analysis of Questionnaire Data. J Med Internet Res 2020 Sep 9;22(9):e21956. [doi: 10.2196/21956]
27. Baptista S, Teles Sampaio E, Heleno B, Azevedo LF, Martins C. Web-Based Versus Usual Care and Other Formats of Decision Aids to Support Prostate Cancer Screening Decisions: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Jun 26;20(6):e228. [doi: 10.2196/jmir.9070]
28. Centers for Disease Control and Prevention. If You Are Sick or Caring for Someone [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/index.html.
29. Microsoft. Microsoft PowerPoint (Version 2019) [Internet]. Redmond, WA: Microsoft; 2021. Available from: https://www.microsoft.com/de-de/microsoft-365/powerpoint
30. InVisisionApp Inc. InVision [Internet]. New York, NY: Microsoft; 2021. Available from: https://www.invisionapp.com/
31. QuestBack GmbH. Unipark [Internet]. QuestBack GmbH; 2021. Available from: ttps://www.unipark.com/
32. Wessel D, Attig C, Franke T. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. Proceedings of Mensch und Computer 2019 [Internet] Hamburg Germany: ACM; 2019 [cited 2021 Aug 26]. p. 147–154. [doi: 10.1145/3340764.3340766]
33. Kim J, Park H-A. Development of a Health Information Technology Acceptance Model Using Consumers' Health Behavior Intention. J Med Internet Res 2012 Oct 1;14(5):e133. [doi: 10.2196/jmir.2143]
34. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist 2003 Mar;38(1):63–71. [doi: 10.1207/S15326985EP3801_8]
35. Kirschner PA, Kirschner F. Mental Effort. In: Seel NM, editor. Encyclopedia of the Sciences of Learning [Internet] Boston, MA: Springer US; 2012 [cited 2021 Aug 26]. p. 2182–2184. [doi: 10.1007/978-1-4419-1428-6_226]
36. O'Connor AM. Validation of a Decisional Conflict Scale. Med Decis Making 1995 Feb;15(1):25–30. [doi: 10.1177/0272989X9501500105]
37. Ammenwerth E. Technology Acceptance Models in Health Informatics: TAM and UTAUT. In: Scott P, de Keizer N, Georgiou A, editors. Applied Interdisciplinary Theory in Health Informatics [Internet] Amsterdam, Netherlands: IOS Press BV; 2019. p. 64–71. Available from: https://ebooks.iospress.nl/publication/51874
38. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]

39. Prolific Academic Ltd. Online Participant Recruitment For Surveys and Market Research [Internet]. 2021 [cited 2021 Apr 9]. Available from: https://www.prolific.co/
40. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. Journal of Experimental Social Psychology 2017 May;70:153–163. [doi: 10.1016/j.jesp.2017.01.006]
41. Rouse SV. A reliability analysis of Mechanical Turk data. Computers in Human Behavior 2015 Feb;43:304–307. [doi: 10.1016/j.chb.2014.11.004]
42. Casey LS, Chandler J, Levine AS, Proctor A, Strolovitch DZ. Intertemporal Differences Among MTurk Workers: Time-Based Sample Variations and Implications for Online Data Collection. SAGE Open 2017 Apr;7(2):215824401771277. [doi: 10.1177/2158244017712774]
43. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021 [cited 2021 May 14]. Available from: https://www.R-project.org/
44. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. Journal of Open Source Software 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
45. Wickham H. ggplot2: 3.3.5 [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021 [cited 2021 May 14]. Available from: https://ggplot2.tidyverse.org/
46. Delacre M, Lakens D, Leys C. Why Psychologists Should by Default Use Welch's *t*-test Instead of Student's *t*-test. rips 2017 Apr 5;30(1):92. [doi: 10.5334/irsp.82]
47. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests [Internet]. 2021 [cited 2021 May 14]. Available from: https://CRAN.R-project.org/package=rstatix
48. Champely S. pwr: Basic Functions for Power Analysis [Internet]. 2020. Available from: https://cran.r-project.org/package=pwr
49. United States Census Bureau. Educational Attainment in the United States: 2020 [Internet]. 2020 [cited 2021 May 2]. Available from: https://www.census.gov/data/tables/2020/demo/educational-attainment/cps-detailed-tables.html
50. Röbbelen A, Schmieding ML, Kopka M, Balzer F, Feufel MA. Data set supplementing "Interactive versus static decision support tools for COVID-19: An experimental comparison" [Internet]. Zenodo; 2021 [cited 2021 Nov 6]. [doi: 10.5281/ZENODO.5651117]

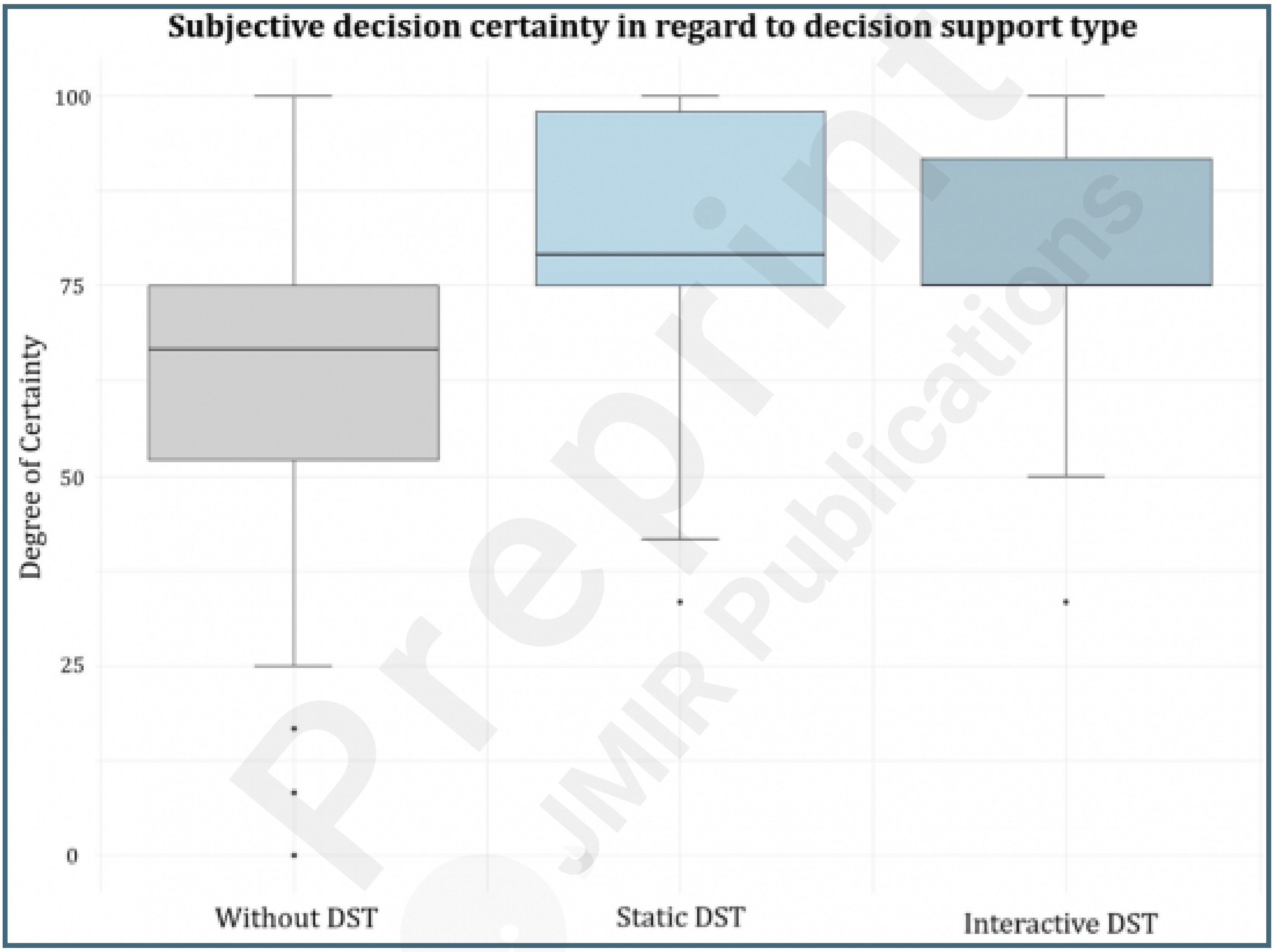
Supplementary Files

Figures

Boxplot showing the distribution of the 196 participants' decision accuracy to appraise 7 fictitious descriptions of patients with symptoms indicative of COVID-19. Study participants (all US inhabitants, non-medically trained, sampled online in November 2020) were tasked to answer two questions per patient description. We randomly assigned participants to one of three experimental groups; in two groups they were supported by either a static decision support tool (DST), ie a flowchart, or an interactive DST, ie a conversational agent mimicking a chatbot. In the control group, they received no decision support. . The boxplots' filled box represents the interquartile range (IQR), the horizontal line inside the box the median, the whiskers the maximum and minimum values within 1.5 IQR of the median, and single dots represent outliers of participants' total number of correct decisions.



Laypersons' perceived certainty in their own appraisals of COVID-19 related clinical decisions obtained in an experimental study in November 2020. The 196 study participants were US residents, non-medically trained and sampled online. Our study tasked them to assess 7 fictitious descriptions of patients with symptoms indicative of COVID-19. Participants were randomized to either receive support from a static decision support tool (DST), ie a flowchart, or an interactive DST, ie a conversational agent mimicking a chatbot. Following the 14 appraisals, we surveyed the participants' certainty in their answers using the Decisional Conflict Scale. A score of 0% indicates minimum certainty, while 100% indicates maximum certainty. The boxplots' filled box represents the interquartile range (IQR), the horizontal line inside the box the median, the whiskers the maximum and minimum values within 1.5 IQR of the median, and single dots represent outliers of participants' total number of correct decisions.



Multimedia Appendixes

Static decision support flowchart for laypersons with suspected COVID-19 infection developed as part of the scientific study.
URL: http://asset.jmir.pub/assets/d89f7b8306d970f5207bfd38be094657.pdf

Interactive decision support tool for laypersons with suspected COVID-19 infection developed as part of the scientific study.
URL: http://asset.jmir.pub/assets/d136f0ac0ec15dc93e06c1c7a4e3b594.pdf

List of COVID-19 symptoms according to Robert Koch Institute.
URL: http://asset.jmir.pub/assets/456747ac2e0694e95ad8ad487bf97ab4.docx

List of risk factors for COVID-19 according to Robert Koch Institute.
URL: http://asset.jmir.pub/assets/d39ef3285d8ca4b584f5da982e9807c3.docx

Overview of the case vignettes developed as part of the scientific study.
URL: http://asset.jmir.pub/assets/5b54b87e4c63076e71ff619e1e80b766.pdf

Outline of all seven case vignettes developed as part of the scientific study.
URL: http://asset.jmir.pub/assets/acb762eee2f948134f46f5a16784dbb0.docx

Multiple choice questions used to assess subjects' prior knowledge.
URL: http://asset.jmir.pub/assets/469d6c6e2acd3b544c7a50beb5acb954.pdf

Decisions on each case vignette in the control group without decision support.
URL: http://asset.jmir.pub/assets/f322d319d0db10a24bc10e661f9c679c.pdf

Decisions on each case vignette in the intervention groups with decision support, here illustrated by the flowchart example.
URL: http://asset.jmir.pub/assets/7f644540075768817eb1cf4fae0b3d6e.pdf

Single Item for Mental Effort.
URL: http://asset.jmir.pub/assets/e9194eb2f709bde9758f536b6e509853.pdf

Decisional Conflict Subscale.
URL: http://asset.jmir.pub/assets/5deea9f67b0b511fbcbe79386a1e925b.pdf

Items measuring participants’ perspectives on the two Decision Support Tools with items on ease of use, usefulness, trust and future intention to use the tool.
URL: http://asset.jmir.pub/assets/7bbcc91cd6868137e787d89046f6b829.pdf

Boxplot for decision accuracy in regard to decision support type. Responses where participants confused "quarantine" and "isolation" were rated as wrong in this analysis.
URL: http://asset.jmir.pub/assets/d4854692c3867566962321c99c19db79.pdf