

that it is both based on economic decision theory and it leaves patient preferences disaggregated from population preferences in an HTA.

Conclusions. The frameworks identified in this review offer potential approaches to systematically and transparently integrate PPs into HTA and decision-making. Based on the review findings, we propose a research agenda to explore the potential of CCA in particular. We anticipate that our findings will augment the recommendations of the Innovative Medicines Initiative PREFER project, which are expected to report in 2022.

OP86 Chatbot-Based Symptom-Checkers: A Systematic Review

Reinhard Jeindl (Reinhard.Jeindl@aihta.at) and
Gregor Goetz

Introduction. Symptom-checkers are digital health applications (DHA) with diagnostic algorithms. These symptom-checkers claim to improve the diagnostic process and patient guidance. After asking the user to describe the symptoms using a chatbot interface, the symptom-checkers offer a list of potential diagnoses, and/or give recommendations for appropriate action (self-care, doctor's visit, or emergency care). Because of the growing number and increasing use of these diagnostic DHA, there is a need to evaluate the evidence.

Methods. We updated a British evidence synthesis on symptom-checkers from the National Institute for Health Research (NIHR, 2019). For the systematic update search, we selected four databases. The following endpoints were selected: effectiveness, safety, diagnostic accuracy, triage accuracy, organizational and patient-relevant endpoints. For accuracy studies included from the update search, we assessed the risk of bias (RoB) using the quality assessment tool of diagnostic accuracy studies (QUADAS-2).

Results. The NIHR-report included 27 studies. We added 14 additional studies via update search. One randomized-controlled-trial (RCT) reported a prolonged illness duration when using symptom-checkers (statistically non-significant). No harms when using symptom-checkers were identified (six observational studies). The diagnostic accuracy ranged from 14-84.3 percent (ten observational studies), the triage accuracy ranged from 33-100 percent (eleven observational studies). For organizational endpoints, the results were inconsistent (one RCT, six observational studies). The patient perspective indicates a high usability for symptom-checkers, but the limited description of symptoms and the missing verbal interaction with health personnel were mentioned as hindering factors (nine survey-studies). The QUADAS-2 assessment for RoB was low in one, and high in seven studies.

Conclusions. The studies were often conducted using fictitious case-vignettes, limiting the validity of the evidence. Therefore, the results for the diagnostic and triage accuracy are insufficient to demonstrate a benefit in real-world settings. Additionally, there is a concern for misdiagnosis and overdiagnosis. We recommend a continuous monitoring of these diagnostic DHA, using high-quality studies.

OP87 Value From A Multi-stakeholder Perspective: A Framework To Assess Digital Health Solutions For Improving Chronic Disease Management

Madeleine Haig (m.haig@lse.ac.uk), Caitlin Main,
Danitza Chavez-Montoya and Panos Kanavos

Introduction. Innovative digital health technologies (DHTs) may present new aspects of value that are not appropriately accounted for in current health technology assessments. In discovering what value means in the context of DHTs, multi-stakeholder collaboration is essential.

Methods. A scoping literature review was conducted to identify current value assessment criteria and proposed methodologies across three health systems: United States of America, United Kingdom, and Germany. A Delphi exercise was conducted with stakeholders from the following groups: users, healthcare practitioners, decision-makers, supply-side actors, and influencers. Based on a review of assessment frameworks in the study countries and consultations with experts from each stakeholder group, researchers proposed value constructs in five domains: health inequalities, data rights and governance, technical and security, economic characteristics, clinical characteristics, and user preferences. In Delphi round one, participants commented on the proposed constructs and submitted their own. A thematic analysis identifying key concepts and themes of the participant proposed constructs and comments was used to incorporate this information for round two. Then, participants rated each value construct on an 'importance' Likert scale in two decision contexts: user-facing DHTs and system-facing DHTs. In round three, participants were presented with the consensus judgement for each construct, with the opportunity to change their answer. Value constructs with equal to or greater than 70 percent consensus were included in the final framework. Rounds four and five were, respectively, value judgements on a Likert scale and a presentation of consensus for a therapeutic area to test the final framework.

Results. Initially 32 value constructs were proposed by researchers, 20 of which were changed or removed based on round one feedback. Additional constructs were added based on participant suggestions resulting in forty-five value constructs in round two. The final framework will be available after round three closes on 20 December 2022.

Conclusions. The multi-stakeholder Delphi approach ensures that all suggestions and value judgements are weighted equally across stakeholder groups. The resultant value framework can be used to inform policymaking around health technology assessment of DHTs.