BMJ Health & Care Informatics

COVID-19 surge readiness: use cases demonstrating how hospitals leveraged digital identity access management for infection control and pandemic response

George A Gellert ,[1] Sean P Kelly,[2] Allen L Hsiao,[3] Brian Herrick,[4] Donna Weis,[5] Jeffrey Lutz,[6] Glynn Stanton,[7] Santos Bonilla,[7] Daniel Borgasano,[8] Matthew Erich,[8] Claire Reilly,[8] Daniel Johnston[9]

To cite: Gellert GA, Kelly SP, Hsiao AL, *et al*. COVID-19 surge readiness: use cases demonstrating how hospitals leveraged digital identity access management for infection control and pandemic response. *BMJ Health Care Inform* 2022;**29**:e100680. doi:10.1136/bmjhci-2022-100680

Received 12 September 2022
Accepted 05 November 2022

For numbered affiliations see end of article.

Correspondence to
Dr George A Gellert;
ggellert33@gmail.com

ABSTRACT

Background Surging volumes of patients with COVID-19 and the high infectiousness of SARS-CoV-2 challenged hospital infection control/safety, staffing, care delivery and operations as few crises have. Imperatives to ensure security of patient information, defend against cybersecurity threats and accurately identify/authenticate patients and staff were undiminished, which fostered creative use cases where hospitals leveraged identity access and management (IAM) technologies to improve infection control and minimise disruption of clinical and administrative workflows.

Methods Working with a leading IAM solution provider, implementation personnel in the USA and UK identified all hospitals/health systems where an innovative use of IAM technology improved facility infection control and pandemic response management. Interviews/communications with hospital clinical informatics leaders collected information describing the use case deployed.

Results Eight innovative/valuable hospital use cases are described: symptom-free attestation by clinicians at shift start; detection of clinician exposure/contact tracing; reporting of clinician temperature checks; inpatient telehealth consults in isolation units; virtual visits between isolated patients and families; touchless single sign-on authentication; secure access enabled for rapid expansion of personnel working remotely; and monitoring of temporary worker attendance.

Discussion No systematic, comprehensive survey of all implemented IAM client sites was conducted, and other use cases may be undetected. A standardised reporting/information sharing vehicle is needed whereby IAM use cases aiding facility pandemic response and infection control can be disseminated.

Conclusions Clinical care, infection control and facility operations were improved using IAM solutions during COVID-19. Facility end-user innovation in how IAM solutions are deployed can improve infection control/patient safety, care delivery and clinical workflows during surges of epidemic infectious diseases.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The innovative and adaptive deployment of identity access and management technology to improve hospital infection control and pandemic response has not been previously reported in the literature.

WHAT THIS STUDY ADDS

⇒ Eight use cases successfully deployed by hospitals in the USA and UK to improve SARS-CoV-2 facility infection control and pandemic response are reported.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ With only two-thirds of humanity currently vaccinated against COVID-19, more virulent, contagious or vaccine-resistant variants of the virus may cause increased community transmission and future surges in patient volume, and these use cases can help hospitals improve their infection control and pandemic operations.

INTRODUCTION

COVID-19 disrupted the already complex digital identity and information environment of modern hospital care delivery and accelerated adoption of telehealth/telemedicine. Hospitals needed to ramp up clinical staff rapidly to manage an increased volume of very ill patients; clinicians and administrative staff had to significantly alter workflows and worksites; and individuals not serving in direct clinical care roles worked remotely, all while maintaining rapid, secure access to critical applications and data. New non-traditional treatment centres—in tents and mobile units, at hotels—were established and had to use existing information technology (IT) to support patient care and information security. Devices used to access information and communicate internally required rapid adaptation to reduce risk of viral transmission within the hospital. Use of mobile devices increased as iPads/tablets were used to support telehealth and facilitate virtual

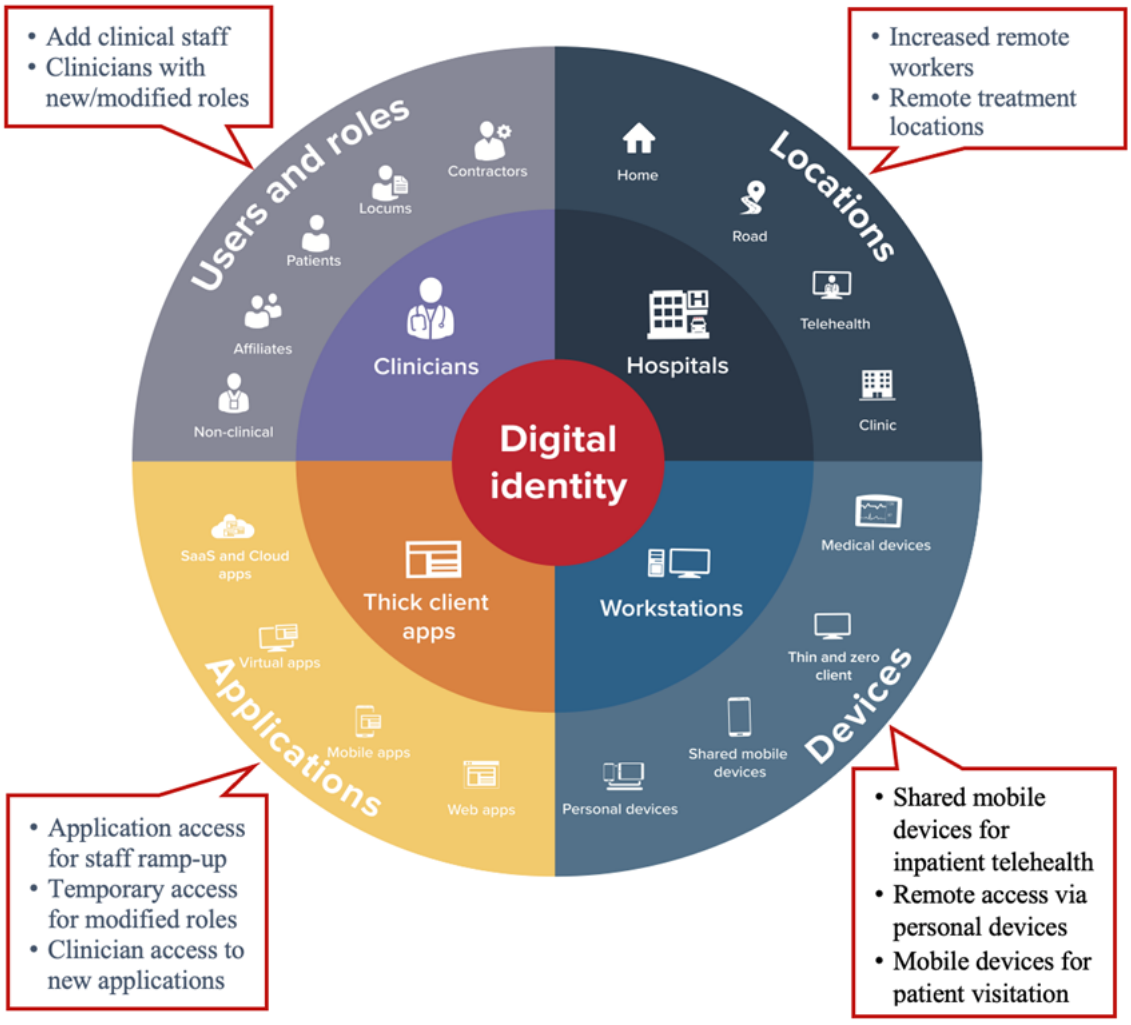


Figure 1 Hospital identity and information access challenges during COVID-19.

patient visits in hospitals, many requiring secure access management and a rethinking of existing high touch processes.

The pandemic surges amplified the centrality of securing and managing digital identity (figure 1). Identity access and management (IAM) capabilities enabled hospitals to leverage these technologies in innovative ways to support their COVID-19 response. A literature review found no use cases reporting IAM technology deployed to improve hospital infection control or outbreak management. Eight use cases are reported here which improved hospital operational and clinical response, reduced potential infection transmission within facilities and helped care providers and administrative staff, as well as patients and their families, cope with the challenges and risks created by the pandemic.

Single sign-on (SSO) expedites use of the electronic health record (EHR) by enabling a clinician to log in by keyboard only once at the start of a shift, and then use a proximity identity badge to reconnect for subsequent logins during the rest of the shift. SSO eliminates need to remember complex passwords, reduces repetitive manual logins and expedites authenticated access to the EHR and clinical software applications. SSO technology liberates substantial time from the keyboard for clinicians to focus on care delivery,[1–4] even more imperative during critical surges in patient volume. IAM remained critical to securing the trusted digital identities of clinicians and patients during the pandemic. Role-based access to quickly on-board clinical and support staff in the face of high patient volumes was imperative. This involved rapidly provisioning application access to accommodate the ramp-up in staff needed to manage high patient volumes and changing the access of certain clinical roles. Enabling access to shared mobile devices for clinician and patient use had to be accomplished securely, accurately and rapidly, as was secure access for a partially remote workforce.

The use cases were deployed to facilitate critical hospital operations during COVID-19 surges, and extend beyond the design intent of the solution vendor in three areas: (1) new workflows to monitor and mitigate risk of viral transmission and hospital-acquired infection within a facility; (2) inpatient telehealth care between care providers, and

virtual visits of isolated patients with family members, to reduce risk of spreading infection within facilities; and (3) enabling IAM during rapid ramp-up of an expanded remote workforce. The technologies deployed in these use cases were Imprivata OneSign for SSO, Confirm ID for clinician identity and multifactor authentication and PatientSecure for patient identity validation.

METHODS

The reported use case data were gathered through communications between the customer support team of a leading IAM solution provider and health IT leaders among its hospital/health system customers. The particular IAM vendor was selected because of its high market penetration in the USA and UK, and because it was facilitating and recording which of its hospital customers were deploying its solutions in innovative ways to improve hospital infection control and COVID-19 operational response. Hospital/health system clinical informatics leaders were contacted in order to solicit a detailed description of how IAM and SSO technologies were leveraged to improve various clinical, infection control and/or operational workflows during the pandemic. Eight use cases deploying IAM technology to improve hospital/health system COVID-19 response were identified, all within the USA and the UK. These nations were a focus because they represent 80.2% (2320 US facilities) and 5.5% (158 UK facilities) of the IAM solution provider's total customer implementations worldwide.

Each use case was documented and shared with the involved hospital facilities in order to validate and improve the accuracy of its description. All recommended facility changes in descriptive use case content were incorporated, and the final report was shared with all hospitals/health systems for final review and approval. All hospitals/health systems known by the vendor to have deployed an innovative IAM use case during the pandemic and contacted in the development of this report also approved their identification, with the exception of one centre which was non-responsive and thus excluded.

RESULTS

Table 1 summarises the hospital facility value and functional focus of each of the eight use cases reported.

Table 1 Use cases of IAM technology deployed in hospital COVID-19 response by value and functional focus

Use case value	Use case functional focus
Infection control and patient safety	SSO enabled clinicians to attest being symptom free at shift start
Infection control and patient safety	SSO deployed for exposure and contact tracing of facility clinicians
Infection control and patient safety	SSO deployed to enable mandatory clinician temperature checks/reporting
Infection control, patient safety and PPE supply chain management	Inpatient telehealth consults and virtual inpatient rounding in isolation units to reduce infection risk and rate of PPE consumption
Infection control and patient/family well-being and psychosocial support	Mobile devices enabled virtual visits between isolated patients and families
Infection control and expedited authentication and workflows	SSO rapidly authenticated into mobile devices without touching screens
Infection control and maintenance of facility organisational effectiveness and work productivity	Secure access enabled for rapid expansion of personnel working remotely
Organisational staffing management, accountability and work productivity	SSO monitored attendance of temporary workers

Key:
IAM, identity access and management; PPE, personal protective equipment; SSO, single sign-on.

SSO enabled clinicians to attest being symptom free at shift start

During an outbreak of a highly transmissible pathogen such as SARS-CoV-2, clinicians can inadvertently spread infection across the hospital. Having clinicians attest at the start of shifts that they were symptom free was critical to reducing viral spread. However, the symptom attestation process must be simple and rapid to ensure compliance and avoid disruption of clinical workflows and care delivery. To enable such rapid attestation and reduce the risk of clinicians adding to the facility burden of infection, hospitals needed a way for clinicians to log in and attest to 'absence of symptoms' with real-time reporting. Employees not providing care needed to be differentiated from those working in a clinical setting with elevated transmission risk to vulnerable patients and clinical colleagues.

A capability for attesting to the absence of COVID-19 symptoms was implemented which did not require all hospital staff but only clinicians at greatest risk of infection transmission to attest. Hospitals leveraged SSO to enable reporting of symptoms among only clinicians through a home-grown survey application that assessed for COVID-19 symptoms according to guidelines iterated by the US Centers for Disease Control and Prevention. This survey function was linked to SSO via an application programming interface so that when a user logged into a workstation, clinicians were automatically prompted to respond. There was no need to enter a username or password to log into symptom attestation, and multifactor authentication enabled users to quickly verify their identity easily and securely.

Hospitals implemented this functionality so it would not be intrusive and would only prompt clinicians once per 12-hour shift at shift start. Clinicians simply tapped their proximity card when prompted in the health attestation, confirmed they were symptom free, and could then begin work. Responses indicating a clinician could be positive for SARS-CoV-2 infection were automatically transmitted in real time to the hospital infection control team for review and appropriate response.

SSO deployed for exposure and contact tracing of facility clinicians

Yale New Haven Health, a system with seven hospital campuses, created a real-time process to monitor

exposure and infection spread by using SSO workstation login records to help track and reduce risk of transmission from potentially exposed individuals. SSO reporting capabilities coupled with location logs enabled identification of exactly where and when users accessed specific workstations across all patient care areas, including those with infection risk. SSO audited the activity of clinical users when users authenticated to workstations, including user identity, workstation and date/time. Audit data retrieved from SSO cross-referenced with the known location of workstations (eg., Nurses Station 3 East) enabled granular infection contact tracing.

Combined with EHR data and workstation mapping, Yale first deployed SSO to accurately track infection exposure and transmission risk for measles and used the same approach to control facility spread of SARS-CoV-2. As clinicians are at risk of contracting infection when treating infected patients and may be exposed before patients exhibit symptoms, SSO detected if clinicians accessed a workstation near a patient or another clinician who subsequently tested positive. The real-time data generated by SSO enabled the facility to identify clinicians who had been in patient care areas where there was high potential risk of exposure and contracting infection.

Yale used multiple data points such as staffing lists to assess risk, while SSO conveyed granularity to identify specific users accessing workstations in units at elevated risk of infection exposure. With SSO providing the date and time of access, and how long it was used (duration of exposure), hospital infection control identified clinicians potentially at risk who accessed a workstation near a patient who was confirmed positive. This was accomplished by analysing data SSO collects when clinicians tap their proximity badge to access a workstation. By matching SSO audit data to location of workstation and patients confirmed positive, hospitals in future surges can determine which specific users were in areas with elevated risk of infection, for how long, and the infection control team can take necessary steps to interrupt further disease transmission. Leveraging SSO data in such a manner can be a powerful tool for infection control teams working to minimise pathogen spread during unexpected or novel outbreaks.

SSO deployed to enable mandatory clinician temperature checks/reporting

During COVID-19 surges, hospitals required clinicians to check and report their temperature twice per shift to monitor for potential infection. Hospitals and health systems with SSO, such as Yale New Haven Health, developed and deployed a home-grown internet-based application to support this workflow that was minimally disruptive for clinicians. Integrated with SSO, clinicians were able to badge tap into a workstation and access the application to report their temperature, ensuring the added workflow was fast and easy to complete. SSO also provided audit data to help hospitals track compliance with the twice per shift temperature reporting requirement. Here again, hospitals identified a use for SSO that was beyond its original design intent, but which delivered critical value in managing the crisis precipitated by COVID-19 volume surges.

Inpatient telehealth consults and virtual inpatient rounding in isolation units to reduce infection risk and consumption of personal protective equipment

During pandemic response, hospitals needed to minimise non-essential in-person interactions between care providers and infected patients to mitigate infection spread and to conserve limited supplies of personal protective equipment. Hospitals such as Nebraska Medicine, and in the UK the Royal Surrey NHS Foundation Trust, used iPads/mobile tablets or smartphones to facilitate on-site telehealth sessions with infected patients in isolation. Hospitals enabled clinicians to conduct clinical televisits without elevating clinician exposure/infection risk. Reducing contact with infected patients was critical to reducing viral exposure and transmission within care delivery settings. Royal Surrey used Ascom smartphones and the Attend Anywhere video consulting application for virtual rounds in the intensive care unit (ICU), with one physician rounding in person while linked remotely to colleagues.

To ensure patient confidentiality, a unique sign-on for each inpatient telehealth encounter was needed, and it was imperative to institute a hands-off login process. A comprehensive mobility solution (Imprivata GroundControl) was implemented at Nebraska Medicine to deliver automated provisioning, secure checkout and fast access to devices and applications. Clinicians tapped their proximity card on a docking station to check out a tablet for their shift. When accessing applications on the device, proxy credentials eliminated clinician need to manually type username/password. This enabled hospitals to automatically provision and digitally sanitise shared tablets, ensuring patient privacy through compliance with the Health Insurance Portability and Accountability Act (HIPAA), and helped hospitals set up, personalise and secure shared tablets. When finished, the device was returned to the docking station to be reset/cleaned and recharged for the next user.

Mobile devices enabled virtual visits between isolated patients and families

During COVID-19, hospitals instituted patient visitation restrictions to reduce risk of spreading infection. Patients—some at risk of death—were isolated from and unable to communicate with loved ones. Enabling virtual visits between isolated patients and family members on shared mobile devices was an important and humane part of care delivery during the COVID-19 crisis. A mobility solution for clinician telehealth encounters with isolated patients also enabled hospitals such as the University of Rochester Medical Center (URMC) to provide patients iPads/tablets for safe family visitation without sharing air space with infected patients.

Mobile devices allowed quick and secure access to conferencing applications, enabling virtual inpatient family visits without risk of infection. Consumer chat applications such as Skype, Facebook Messenger and FaceTime were deployed with requisite privacy and information security measures. In providing iPads to patients for communication with visitors, URMC needed a way to provision, configure, wipe clean and reset devices back to a ready state for next patient use. In addition to physical disinfection, with SSO and a mobility solution it was possible to digitally sanitise identity on shared iPads/tablets while they were recharged to set up and secure the devices for use by other patients. In the UK, the Royal Surrey NHS Foundation Trust also provided virtual visits for patients in isolation, repurposing shared mobile devices accessed through SSO. A link sent to the patient's family connected to the virtual meeting with the patient, and at session end patient access was tapped out and the device cleared and cleaned for next patient use.

SSO rapidly authenticated into mobile devices without touching screens

During the pandemic inpatient nurses used mobile devices to share real-time patient vitals with remote physicians (e.g., in the ICU). Manual login while gloved is difficult and consumes time. Sharing mobile tablets increases risk of spreading infection, and during surges hospitals used SSO to enable staff to tap their badge to log in and out of devices during clinical care delivery. The Cambridge Health Alliance deployed SSO into their ICU Microsoft Surface Pro tablets to log in simply and rapidly and connect with remote physicians. SSO enabled badge tap in for instantaneous touchless access, mitigating infection risk and facilitating real-time communication with remote physicians.

Secure access enabled for rapid expansion of personnel working remotely

Patient volume surges increased the risk of hospital-acquired infection among attending clinical but also administrative personnel not involved in care delivery, and hospitals rapidly expanded the number of workers shifted to remote work. New York City Health+Hospitals (NYCH+H) sought to enable some clinical personnel, such as consultants, to work remotely when feasible. Remote workers needed network access that protected confidential patient medical information and secured against unauthorised access by cybercriminals taking advantage of COVID-19 to perpetrate phishing attacks.[2] Multifactor authentication was needed to help hospitals ensure information security as their remote workforce expanded. A solution providing fast, secure multifactor authentication for remote access, Imprivata Confirm ID, was provided to hospitals such as NYCH+H and Children's Hospital of the King's Daughters in Virginia. It enabled rapid expansion of secure access for remote workers with minimal disruption of critical workflows. Coventry and Warwickshire Partnership Trust in the UK was able to deliver secure home/remote work for more than 4000 staff from 60 different locations in several days.

SSO monitored attendance of temporary workers

COVID-19 surges forced many hospitals to rapidly ramp up temporary clinical and other staff. A challenge hospitals faced was a need to enable secure access and monitor staffing and work attendance of temporary workers to ensure appropriate staffing needs were met and to enable related financial processes. NYCH+H deployed SSO to enable temporary workers to badge log in at the beginning of shift, just as it does for permanent staff members. SSO authentication enabled hospitals to confirm when temporary workers had started their shift. This eliminated need for additional solutions to address this requirement, alleviating an operational concern during the pandemic when temporary workforces were alternately expanding rapidly and subsequently contracting as patient volume surged and waned.

DISCUSSION

A limitation of the methods used is the very high specificity but low sensitivity of vendor-identified use cases. We did not conduct a systematic survey of possible IAM-enabled COVID-19 response use cases among all customers of this leading IAM solution provider. Thus, it is possible that valuable use cases where hospitals deployed IAM technology to improve pandemic response and infection control were not known to the vendor among thousands of implemented facilities worldwide. Working with IAM solution vendors, researchers can establish a standardised use case surveillance and reporting process and vehicle where such valuable applications of IAM technology can be reported and detailed for sharing with the broader community of hospitals.

Hospitals may deploy these and other innovative applications of existing IT capabilities in future surges of highly communicable diseases, including and beyond COVID-19. Customisation of existing vendor technologies is often more easily and rapidly scalable—especially important during time-critical emergencies such as communicable disease outbreaks. Collaborative partnerships between hospitals and their IT vendors can help facilities implement such innovative solutions, whether in crisis response or routine operations. These use cases suggest hospitals and health IT vendors should regard solutions as an 'innovation sandbox' through which care delivery organisations can explore and innovate needed functionality, adding value and impact to existing products/services. In today's cost-conscious performance-focused healthcare environment—one likely to be challenged recurrently with future care crises—this may increasingly become imperative, not optional.

Hospital clinical, IT and administrative leaders should not regard the products/services they purchase from health IT vendors as static and delimited in terms of problems they can resolve and challenges vendors can help

them meet. The necessity and urgency created by the pandemic crisis fuelled these hospital-inspired innovative applications of SSO and IAM technologies. However, patients, clinicians, administrators and payers alike will benefit when inventive and need-driven creative collaboration between hospitals and their IT vendors becomes the rule, rather than the exception.

CONCLUSIONS

Care delivery, patient-visitor and staff infection control and safety, and facility pandemic operations were improved by hospitals deploying existing IAM solutions creatively during COVID-19 surges. Facility end-user innovation in how IAM solutions are deployed, as driven by need, can reduce hospital spread of infectious pathogens and improve patient safety and care delivery by enabling more effective and safer clinical workflows during surges of highly contagious/epidemic infectious diseases. With only two-thirds of humanity vaccinated against SARS-COV-2, future variants of even greater communicability, virulence and potential vaccine evasion than we have witnessed thus far are possible.[5] Should substantial surges of patients requiring hospital care occur in coming years, facilities should consider whether these (and other) IAM use cases can improve their infection control capabilities and overall hospital COVID-19 response effectiveness.

Author affiliations
[1]Medical Advisor, Impact Demonstration, Imprivata, San Antonio, Texas, USA
[2]Department of Emergency Medicine, Harvard Medical School, Boston, Massachusetts, USA
[3]Department of Pediatric Emergency Medicine, Yale School of Medicine and Yale New Haven Health System, New Haven, Connecticut, USA
[4]Department of Health Informatics, Tufts University School of Medicine, Boston, Massachusetts, USA
[5]Department of Health Informatics, Nebraska Medicine, Omaha, Nebraska, USA
[6]Department of Health Informatics, New York City Health and Hospitals, New York, New York, USA
[7]Department of Health Informatics, Yale New Haven Health System, New Haven, Connecticut, USA
[8]Department of Clinical Operatons, Imprivata, Lexington, Massachusetts, USA
[9]Department of Clinical Operatons, Imprivata UK, Uxbridge, UK

Acknowledgements The authors are grateful to the clinicians who served selflessly during the pandemic, and to the hospital leadership teams whose innovations we have described.

Contributors All coauthors contributed either to the design or implementation of the innovative IAM use cases described, and/or the writing of the resulting manuscript. GAG is the author serving as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests GAG is an external medical advisor to Imprivata.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. All data relevant to the study are included in the article, in the form of output of communications and interviews with hospital personnel. However, no data set was generated for this study.

ORCID iD
George A Gellert http://orcid.org/0000-0002-3519-7486

REFERENCES

1 Gellert GA, Crouch JF, Gibson LA, *et al*. An evaluation of the clinical and financial value of work station single sign-on in 19 hospitals. *Perspect Health Inf Manag* 2019;16:1.
2 Tidy J. How hackers are preying on fears of COVID-19. British Broadcasting Corporation news, March 13, 2020. Coronavirus: how hackers are preying on fears of Covid-19 BBC News; 2022 [Accessed 16 Oct 2022].
3 Griffith A. Eliminate login nightmares with single sign-on technology Health IT Outcomes; 2015.
4 Fontaine J, Zheng K, Van De Ven C, *et al*. Evaluation of a proximity card authentication system for health care settings. *Int J Med Inform* 2016;92:1–7.
5 World Health Organization, Our World in Data, Coronavirus (COVID-19) vaccinations. Coronavirus (COVID-19) Vaccinations - Our World in Data; 2022 [Accessed 16 Oct 2022].

BMJ Health & Care Informatics

Understanding barriers of receiving short message service appointment reminders across African regions: a systematic review

Addisalem Workie Demsash ,[1] Masresha Derese Tegegne ,[2] Agmasie Damtew Walle,[1] Sisay Maru Wubante [2]

To cite: Demsash AW, Tegegne MD, Walle AD, *et al*. Understanding barriers of receiving short message service appointment reminders across African regions: a systematic review. *BMJ Health Care Inform* 2022;**29**:e100671. doi:10.1136/bmjhci-2022-100671

Received 26 August 2022
Accepted 08 November 2022

[1]College of Health Science, Health Informatics Department, Mettu University, Mettu, Ethiopia
[2]Health Informatics Department, University of Gondar College of Medicine and Health Sciences, Gondar, Ethiopia

Correspondence to
Addisalem Workie Demsash; addisalemworkie599@gmail.com

ABSTRACT

Objective Patients frequently miss their medical appointments. Therefore, short message service (SMS) has been used as a strategy for medical and healthcare service appointment reminders. This systematic review aimed to identify barriers to SMS appointment reminders across African regions.

Methods PubMed, Google Scholar, Semantic Scholar and Web of Science were used for searching, and hand searching was done. Original studies written in English, conducted in Africa, and published since 1 December 2018, were included. The standard quality assessment checklist was used for the quality appraisal of the included studies. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart diagram was used for study selection and screening, and any disagreements were resolved via discussions.

Results A total of 955 articles were searched, 521 studies were removed due to duplication and 105 studies were assessed for eligibility. Consequently, nine studies met the inclusion criteria. Five out of nine included studies were done by randomised control trials. The barriers that hampered patients, mothers and other parental figures of children when they were notified via SMS of medical and health services were identified. Among the 11 identified barriers, illiteracy, issues of confidentiality, familiarised text messages, inadequate information communication technology infrastructure, being a rural resident and loss of mobile phones occurred in at least two studies.

Conclusions SMS is an effective and widely accepted appointment reminder tool. However, it is hampered by numerous barriers. Hence, we gathered summarised information about users' barriers to SMS-based appointment reminders. Therefore, stakeholders should address existing identified barriers for better Mhealth interventions.

PROSPERO registration number CRD42022296559.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ In developing regions, patients miss their medical appointment schedules and are less likely to use health services. Though short message service (SMS) has been effective in reminding patients of their medical appointments, it is not yet optimal and not fully functioning.

WHAT THIS STUDY ADDS

⇒ This study incorporates comprehensive information about barriers to SMS-based medical appointment reminders. We indicated possible solutions to fix the identified barriers. We presented the impact of SMS on appointment reminders according to the reports of each study.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study provides good insight into readers about the barriers to SMS. To overcome the identified barriers and adopt digital technology in the healthcare system, stakeholders would use this reviewed work as a state-of-the-art synthesis and the best source of evidence.

⇒ Researchers may use this evidence as input for future research on mobile-based health intervention, remote patient counselling and home-based care.

INTRODUCTION

Patients frequently miss and arrive late for their scheduled medical appointments. Missing medical appointments can be caused by several factors, including forgetfulness, confusion, miscommunication on medical appointment details, feeling better and difficulty in tracking appointment schedules.[1] For instance, of the total medical appointments, more than 3 out of 10 medical appointments are missed by patients.[2] Access to healthcare services and medication is significantly hampered by missing medical appointments. Therefore, patients who regularly miss their appointments use healthcare services less[3] and have poor health outcomes.[4] In Africa, the problem is prevalent.[5] Therefore, mobile technologies and telecommunications have been integrated into the healthcare industry as promising tools for bridging communication gaps, fostering behavioural change, overcoming

missed medical appointments and other challenges that developing regions face.[6]

A short message service (SMS) is used as the medical appointment reminder method for better healthcare outcomes, enhancing appointment attendance[7] mitigating missed medical appointments[8] and treating individuals with health and psychological problems.[9] SMS is a convenient mode of communication that is widely used and is an effective strategy to reach a large population.[10] However, the implementation, adoption, scale-up and sustainability of mobile-based appointment reminders are challenging.[11] SMS-based appointment reminders disintegrated with other digital technology (healthcare system application software) in resource-limited settings.[12] SMS-based appointment reminders are not yet optimal. Patients rarely and never reschedule their medical appointments.[13] Though SMS provides instant and asynchronous communication, it is not standardised and fully functional across different health institutions.[14]

The rationale of the study

Mobile technology utilisation in the healthcare system is inadequately limited to specific functions for health interventions in the developing world compared with other developed regions. Conclusive evidence for health policymakers and professionals is required. Reliable guidelines are developed through research and evaluation to implement mobile-based health interventions and to tackle the possible challenges. However, studies of barriers to SMS-based appointment reminders across African regions were inadequate, and previous work was limited to a study setting. Previously reviewed work has not synthesised evidence on the significant barriers that affect patients, mothers or other parental figures of children to SMS-based appointment reminders. This study provides comprehensive information on the objectives of this study by critically evaluating and synthesising existing primary studies. Therefore, this study aimed to identify patients', mothers' or other parental figures of children's barriers to SMS-based medical or health service appointment reminders. The objectives of this study were: (1) to identify patients', mothers' or other parental figures of children's barriers to SMS-based appointment reminders, (2) to create an understanding of the identified barriers and provide alternative solutions to fix barriers.

METHODS

Protocol and registration

The systematic review protocol was registered in the PROSPERO database. Retrieved from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022296559.

Study design

A systematic literature review was conducted on different articles in the African regions and published in peer-reviewed journals. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[15] were used to map the number of records identified, included and excluded in the review process.

Inclusion criteria

Studies that meet the following inclusion criteria were included for review:

Types of study

Original studies that were conducted based on randomized control trial (RCT), experimental and quasi-experimental, observational, cross-sectional and qualitative research designs, written in English, and peer-reviewed were included. Studies that report barriers to SMS-based appointment reminders among patients, mothers or other parental figures of children were included.

Type of study participants

Studies that examined SMS appointment reminders on any patients, mothers or other parental figures of children who were notified via SMS of their medical, maternal and child care services were eligible in the present systematic review.

Period of publication

To identify the most up-to-date evidence and as we had limited time for database search, articles published since 1 December 2018 were included in this systematic review.

Study areas

Studies conducted across African regions were included.

Intervention

SMS text message-based appointment reminders.

Comparator

Usual patients' appointment reminder systems.

Outcome

The primary outcome was the effect of SMS or text messaging for appointment reminders. The secondary outcome was appointment attendance without SMS or text messaging.

Exclusion criteria

Types of studies

Duplicated studies in preprint, conference and reviewed papers on text message appointment reminders, books, diaries, commentaries and letters about text message appointment reminders; and studies that do not report barriers to SMS-based appointment reminders were excluded from this systematic review. In addition, since we did not have the financial and human resources to translate, papers written in non-English languages were also excluded.

Period of publication

Studies published before 1 December 2018 were excluded due to the limited time we had, and we believed that

Table 1 Synonyms

Keyword	Receive	SMS text message	Reminders	Patients
Synonyms	► Acceptance ► Uptake ► Approval	► Text message ► Short text message ► Short message service	► Memorandum ► Notification	► Mothers

SMS, short message service.

studies published since 1 December 2018 provide up-to-date information.

Study areas

Study participants might not face the same barriers in receiving SMS-based appointment reminders. Therefore, studies conducted outside African regions were excluded.

Information source and search strategy

An electronic database search technique was used to identify peer-reviewed articles. PubMed, Google Scholar, Semantic Scholar and Web of Science were used for study searching. Gray literature on other relevant internet engines and hand searching using the references of the included studies were conducted. The studies, published until 28 February 2022, were identified by searching these electronic databases following systematic review searching procedures. Snowballing was used to look through the references of recognised publications for studies that might be relevant. For studies that might be pertinent, snowballing was employed to search through the references of reputable publications. Possible search words (synonyms) and mesh terms for each keyword were defined (table 1, box 1). Each keyword or term was searched using Boolean operators (OR, AND) in combinations. Specifically, (Receive OR Acceptance OR Approval OR Uptake) AND (Short message services OR text messages OR Short text messages) AND (Reminders OR Memorandums OR Notifications) AND (Patients OR Mothers). The Mesh terms of the PubMed database are presented in table 2.

Condition/domain studied

For this study, SMS is defined as a text message sent to a mobile phone, which includes any Mhealth interventions performed for patients, mothers or other parental figures of children either manually or automatically. Therefore, this systematic review considered original studies conducted on SMS-based appointment reminders for patients, mothers or other parental figures of children as a domain to be explored.

Box 1 The search strategy

A search strategy for PubMed

(((((((((Receiving [(All Fields]) OR Acceptance [(All Fields])) OR Uptake [(All Fields])) OR Approval [(All Fields])) AND

("text messaging"[(MeSH Terms]) OR Short message service [(Text Word])))) OR ("text messaging"[(MeSH Terms]) OR Text Message [(Text Word])))) OR (Short [(All Fields]) AND ("text messaging"[(MeSH Terms]) OR text message [(Text Word]))))) AND

Reminder [(All Fields])) OR Notification [(All Fields])) AND

("patients"[(MeSH Terms]) OR Patient [(Text Word])))) OR ("mothers"[(MeSH Terms]) OR Mother [(Text Word]))

Review and selection process

A Preferred Reporting Item for Systematic Review and Meta-Analysis Protocol was used for study selection and screening.[15] A total of 955 studies were searched. The search results were exported to Endnote V.X9 software, and 521 duplicates were removed. Then studies were examined according to the selection criteria. Titles and abstracts were independently examined for the probability of eligibility by authors (AWD and MDT). Accordingly, a total of 324 studies were removed for a reason. At this stage, disagreements were resolved through discussion. After discussion, six studies were removed since titles and abstracts did not provide adequate information to decide, and ambiguities and uncertainties happened. A total of 104 articles had undergone full-text screening. In the second stage, full-text screening was done independently by authors (AWD and ADW). As a result, 67 articles were removed. Discrepancies between authors were resolved through discussion, and so 18 studies were removed since they dealt with an automated text-generating system. Finally, nine studies were included and reviewed to identify barriers to SMS-based appointment reminders (figure 1).

Data extraction and analysis

Data extractions were done on the included studies. The data extraction process was also done by authors (AWD and SMW) independently using predefined criteria. Discrepancies between authors were resolved through discussion. The studies' characteristics, such as authors name, publication year, study design, data collection methods, study participants, sample size and sampling technique, were extracted independently to describe the included studies. The impacts of SMS and mobile technology on health interventions were assessed according to the included study report. Content analysis was used to group barriers based on their concept of relationship.

Quality assessment and appraisal

Quality assessment criteria were established for studies that reported barriers to SMS-based appointment reminders. Joanna Briggs Institute quality assessment checklists were used to appraise the quality of the included studies.[16]

Table 2 Characteristics of included studies of SMS-based appointment reminders among patients

Author/publication year/country	Study participant/sample size	Study design/methods of data collection/sampling method	Barriers to SMS-based appointment reminders	The impacts of SMS according to the objective of each study
Geldof/2020/Burkina Faso[17]	People living with HIV/3800	RCT/semi-structured interviews and focus groups/purposive	▶ Insufficient ICT literacy skill ▶ Linguistic diversity ▶ Unable to read the messages	Examine SMS to improve antiretroviral (ARV) retention and adherence
Nhavoto/2018/Mozambique[18]	HIV, TV patients+HCWs/141+40=181	RCT/semi-structured interview/ not stated	▶ Matters of confidentiality ▶ Loss of phone (Stolen)	Use of SMS for reminding appointments, medications, and sending motivational texts
Salihu/2019/Nigeria[19]	Patients/200	RCT/questionnaire/random	▶ Being a rural resident ▶ Lack of money ▶ Being of none educated ▶ Lack of transport, ▶ Work and time constraints ▶ Poor awareness of the disease	Role of SMS reminders on the uptake of glaucoma screening
Schroeder/ 2021/Tanzania[23]	A child with cancer/40	Cross-sectional survey/questionnaire/ purposive sampling	▶ Low literacy ▶ Being a rural resident	Assessment of mobile phone ownership, use, and acceptability of an m-health intervention
Adewuya/2019/Nigeria[20]	Adults with depression/895	RCT//stratified random sampling	▶ Lack of smartphone	Effectiveness and acceptability of mobile telephone adherence support for the management of depression
Mekonnen/ 2021/Ethiopia[24]	Mothers/23	Qualitative/interview/purposive	▶ Low mobile phone ownership ▶ A mobile network, ▶ Electricity, Illiteracy ▶ Inadequate ICT infrastructure	Explore acceptability, barriers and facilitators' text message reminders system
Moodley/2019/South Africa[25]	Cervical cancer precursor/364	Mixed method/in-depth interview, FGD and questionnaire/stratified and purposive	▶ Confidentiality ▶ Loss/theft of mobile phones, ▶ Lack of clarity of language ▶ Low education, literacy	Feasibility of m-health technology to improve management and follow-up of clients
Dissieka/2019/Cote d'Ivoire[21]	Mothers/1596	RCT/questionnaire	▶ Lack of familiarity with text messaging ▶ Challenges of reading and understanding written messages ▶ Low levels of education and literacy	Assess the effect of mobile voice or SMS reminders messages on health facility attendance
Endebu/2019/Ethiopia[22]	HIV clients/420	Cross-sectional, qualitative/ questionnaire and FGD/random	▶ Confidentiality	Acceptability and feasibility of SMS to improve ART medication adherence

FGD, Focus Group Discussion; ICT, Information Communication Technology; RCT, randomised control trial; SMS, short message service.

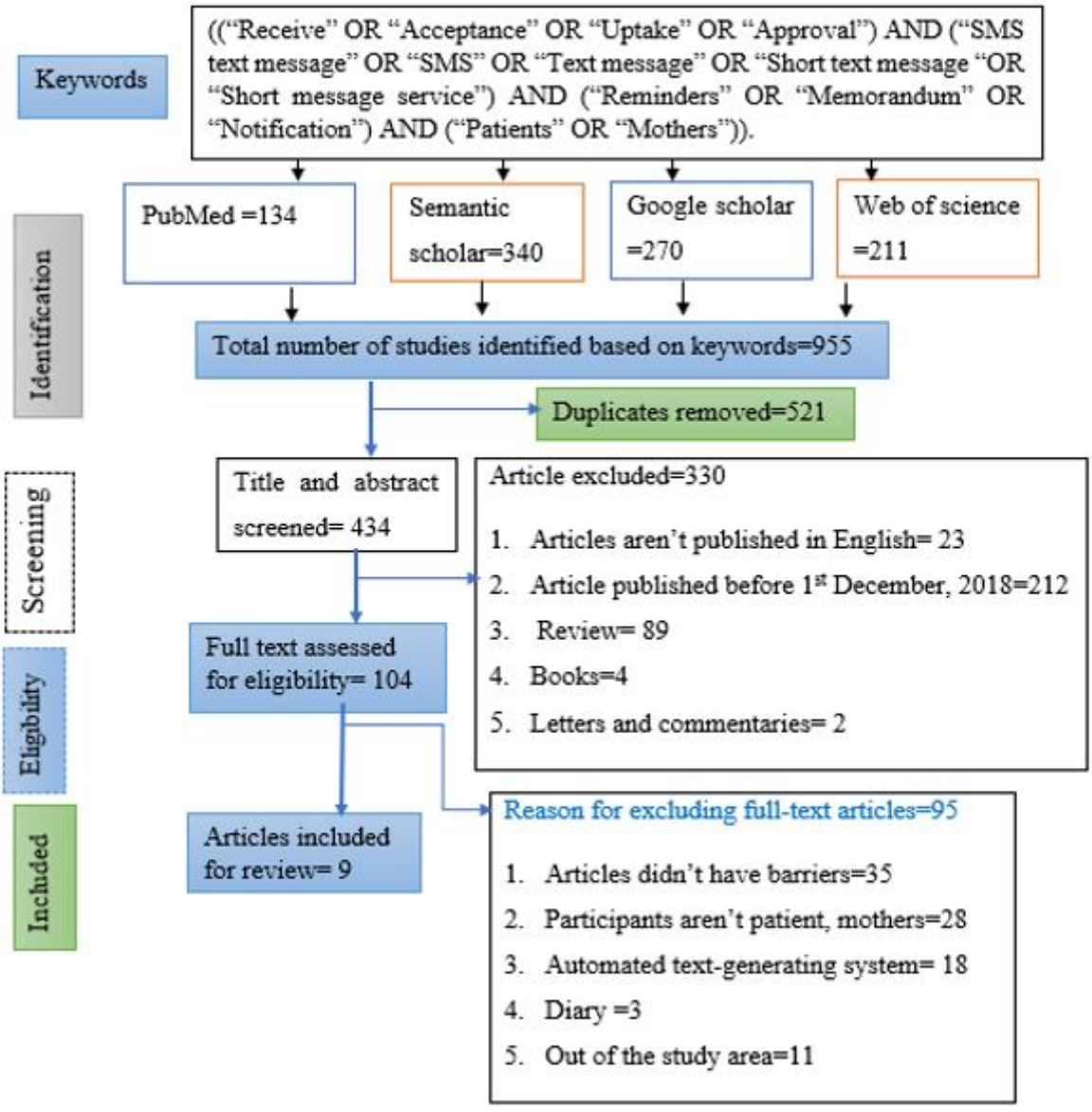


Figure 1 PRISMA flowchart diagram of the studies' selection and screening. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Issues that led to possible bias that occurred at all stages of the review were reduced and addressed. For instance, to address location and selection bias, studies published in an international electronic database were retrieved, and authors (AWD and MDT) assessed the quality of the studies independently. The authors met periodically to discuss biased concepts, and a consensus was reached according to the authors' discussion. There is no validation of the agreement between the authors for each disagreement.

RESULTS

A total of 955 studies were identified through search strategies. From a total of 955 studies, 521 were removed due to duplication. The remaining 434 studies were screened for further processing, and 329 were excluded after reviewing the titles and abstracts due to being unable to meet the inclusion criteria. Then, 105 full-text articles were assessed for eligibility based on the predefined criteria, and 95 articles were excluded for a reason. Overall, nine studies met the eligibility criteria and were included in this systematic review (figure 1).

Features of the included studies

Of the total of 955 identified studies, nine studies were selected and reviewed for the synthesis of evidence after excluding ineligible studies. The study design of the included articles was RCT, qualitative, mixed and cross-sectional. Five out of nine studies (55.6%) were done by RCT.[17–21] Two out of nine studies (22.2%) were done with a cross-sectional study design,[22 23] whereas the remaining two studies (22.2%) were done by qualitative methods[24] and a mixed method[25] of study design accounted for one in each study design. Four out of nine studies (44.4%) were done in Nigeria[19 20] and Ethiopia[22 24] accounted

for two studies in each country. The remaining five out of nine studies (55.6%) were done in Burkina Faso,[17] Mozambique,[18] Tanzania,[23] South Africa[25] and Cote d'Ivoire[21] with one in each country. From the quantitative studies included, the maximum sample size was 3800, which was done in Cote d'Ivoire to assess the effect of SMS on health facility attendance among people living with HIV.[21] According to the publication year, five out of nine (55.6%) studies were published in 2019.[19–22 25] Two out of nine studies (22.2%) were published in 2021,[23 24] and the remaining two out of nine studies (22.2%) were published in 2018,[18] and 2020[17] accounting for one for each year (table 2).

The impact of SMS and mobile technology on health interventions

According to included studies report, SMS and mobile technology had an impact on improving antiretroviral retention and adherence,[17] sending motivational texts,[18] assessment of mobile phone ownership and use of SMS for health interventions,[23] effectiveness, feasibility and acceptability of mobile telephones for adherence support, depression management.[25] SMS-based reminder messages had a positive effect on health facility attendance and uptake of health services.[21] The studies suggest that SMS-based text message is an acceptable mobile health strategy for health interventions and it is feasible, and effective in terms of scheduling appointments, receiving follow-up messages and health facility visits, for appointment cancellations, communication and information sharing (table 2).

Identified barriers for patients and mothers while they receive SMS-based appointment reminder messages

As presented in table 2, barriers to patients, mothers or other parental figures of children while they receive SMS-based medical or healthcare service appointment reminders were identified across nine included studies. Though content analysis is usually used for qualitative research studies, the identified barriers were grouped based on their concept of relationship. Overall, 11 barriers were identified based on their concepts of relationship across the nine included studies.

Of these, illiteracy was one of the barriers to SMS-based appointment reminders, as reported by six studies.[17 19 21 23–25] In this systematic review, patients' inability to use mobile phones or text message reminders, poor awareness and knowledge of the disease, being unable to respond with SMS message information, having no or low-level education, being unable to read and write the SMS message and being unable to understand and comprehend SMS messages were grouped as illiteracy. So, illiteracy was the first most common barrier, occurring six times out of the nine included studies (54.6%). Confidentiality issues[18 22 25] were the second most common barrier, occurring three times out of the nine included studies (27.3%). In this study, confidentiality was understood as the intentional disclosure and access of health information and mobile phones without the users' consent, including privacy and security issues.

Inadequate ICT infrastructure[17 24] including poor access to electricity and network; unfamiliarity with text messages,[21 25] including unclear language, impersonal nature and negative, meaningless and uninterpretable SMS messages; and being a rural resident[19 23] were the third most common barriers, occurring two times out of the nine included studies (18.18%). Unfamiliarised (with) SMS messages in this systematic review means the negative and impersonal nature of messages, SMS messages that are inadequate to add insights, and uninterpretable messages. Loss of mobile phones in this systematic review included loss of mobile phones, and their function as well as being stolen.[18 25] Low mobile phone ownership,[24] linguistic differences, that is, the presence of different languages in the country,[17] lack of money, lack of transport, work and time constraints[26] were the least frequent (9.1%) barriers, occurring once out of the nine included studies (figure 2).

DISCUSSION

In this systematic review, original studies focused on SMS-based medical and health service appointment reminders for any Mhealth intervention among patients, mothers or other parental figures of children across African regions were included for review. Accordingly, 9 out of the 955 searched studies were included and reviewed. A total of 11 barriers have been identified that could potentially hinder patients, and mothers from receiving SMS-based appointment reminders.

Of those barriers, illiteracy was the first main barrier to patients, mothers or other parental figures of children receiving SMS-based appointment reminders, which occurred six times out of the nine included studies.[17 19 21 23–25] This evidence is the same as a study report that highlights illiteracy among a large population as a barrier to mobile health wallet implementations.[27] This barrier might occur when literacy rates and educational status are low.[28] Illiteracy in this systematic review incorporates patients' inability to use mobile phones or text message reminders, insufficient skill, poor awareness and knowledge of the disease and being unable to operate mobile phones and respond with SMS message information. Therefore, enhancing the educational status of mobile phone users, sending SMS messages based on the receiver's preferences and giving them training about how they use and operate mobile phones might be solutions.

Three studies reported that confidentiality was highlighted as one of the second-most common barriers to SMS-based appointment reminders.[18 22 25] This finding is similar to a review report.[29] Although it may be an unintentional disclosure of an individual's health status, the study participants (patients with HIV and TB) were concerned and feared unauthorised access to their text messages.[21] The finding is also similar to health

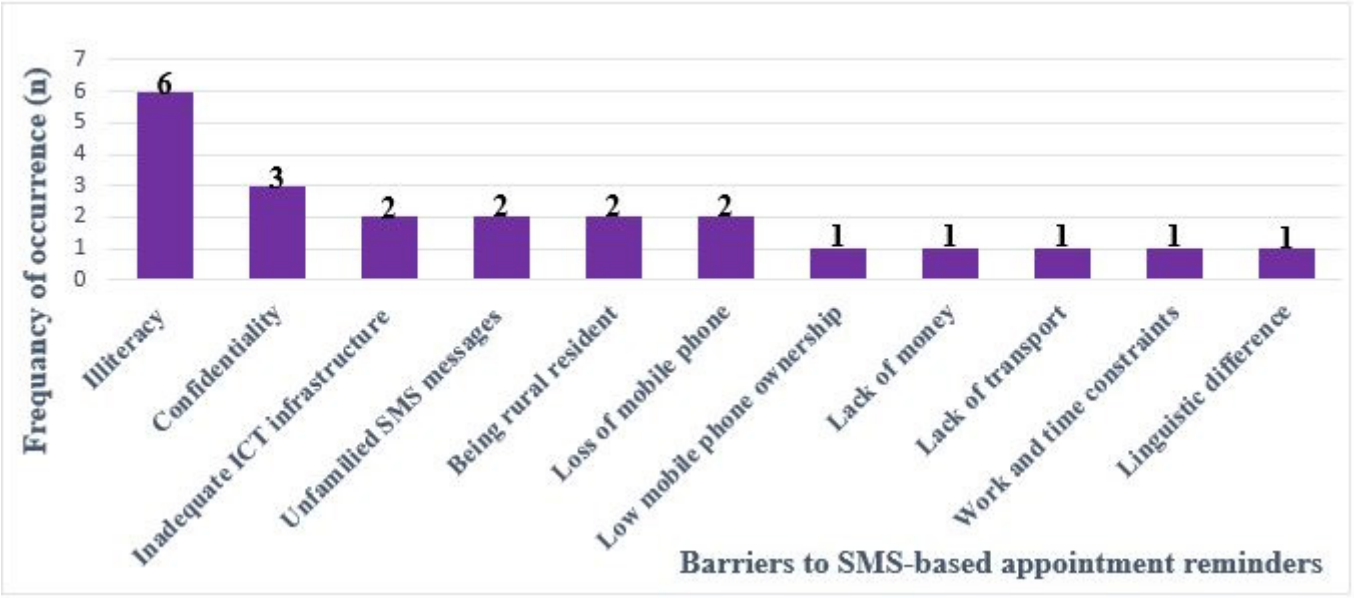


Figure 2 The frequency of occurrence of identified barriers to SMS-based appointment reminders. SMS, short message service.

professionals' confirmations that communicating via SMS can breach the confidentiality of patient information when SMS messages are sent to patients.[25] A patient with HIV stated that receiving SMS text messages disseminates and discloses health information to others.[22] This is because of using shared phones to receive SMS reminders, accessing one's phone without consent; leaving a mobile phone with a neighbour to charge the battery,[18 22 25] and unclear privacy and security regulations.[12] Untrained users may find it difficult to deal with smartphone technology. Unsecure wireless networks and leave the mobile phone where others can access and read SMS messages; health information may also be at risk if stored in a non-secure location.[30] As a result, confidentiality issues make many patients abandon and give up their treatment, and they are less likely to receive SMS-based appointment reminders. Under this systematic review, confidentiality issues include the intentional disclosure of health information, privacy and security problems and accessing information and mobile phones without the users' consent. So, an agreed code and a real-time communication system would maintain patients' confidentiality issues.

Unfamiliarity with text messages, insufficient ICT infrastructure, being a rural resident and phone losses were significant barriers to SMS-based appointment reminders. Unfamiliarised text messages were identified as a barrier according to two studies' reports.[21 25] In this systematic review, unfamiliar SMS messages indicate a lack of language clarity, negative messages, inadequate to add insights and uninterpretable SMS messages. Hence, patients are challenged to read and understand the SMS text messages and face a problem communicating with a service provider about the timing of facility visits.[21] Therefore, sending clear SMS messages that would enhance patients' insight, having oral communication (phone calls) with patients, developing mobile apps for visual communication and sending code and image-based reminder systems are preferred to overcome such barriers.

Inadequate ICT infrastructures like poor access to networks and electricity were identified as barriers to SMS-based appointment reminders.[17 24] A study has also shown that unreliable ICT infrastructure and a shortage of apps and hardware devices are challenges for Mhealth implementation.[31] Though ICT infrastructures are prerequisites for advanced technology, lack of ICT infrastructure coverage and inaccessibility of technologies such as reliable networks, the internet and electricity access are the main challenges to SMS-based appointment reminders.[32 33] We recommend that stakeholders instal a reliable network that covers large geographic areas with adequate and reliable Information Communication Technology (ICT) infrastructure to support mobile phone users and enhance mobile-based health interventions.

Two studies reported that being a rural resident is highlighted as a barrier to SMS-based text message medical appointment reminders.[19 23] Since infrastructure in rural communities is the main challenge, mobile networks and electric power in rural communities are less accessible. So, the government should instal networks and electric power for rural communities and expand 4G and 5G coverage in these areas. The loss of a mobile phone is a barrier to SMS-based medical appointment reminders.[18 25] Even if 98% of women with cervical cancer own mobile phones, only 50% of them attend their appointments, and 58% of them miss their appointments due to theft or loss of a mobile phone.[25] In this systematic review, the loss of a mobile phone means the loss of a mobile and its function as well as being stolen by another person. We recommend

that mobile phone users take care to prevent theft and protect their mobile phones and SIM cards from failure and loss.

In this systematic review, lack of money is identified as a barrier to SMS-based appointment reminders. Lack of money includes absence and insufficient birr, resource, financial and mobile card problems. This happens when there are financial problems, the high cost of mobile phone ownership and operation and the low amount of a person's income necessary to connect to technology.[26] So, supporting users in cash and with a free mobile package of services. Lack of transport, work and time constraints were barriers for patients, mothers or other parental figures of children to receive SMS-based appointment reminders. As clearly presented in the reviewed article, being too far from the hospital and having a long travel time were the constraints for the patients, which explained why they failed to attend their medical appointment.[19] This might be associated with how often and when receivers are notified via SMS.[29] Therefore, reducing travel distance and travel time by constructing nearby health facilities and giving free transport services will increase patients', mothers' or other parental figures' attendance at the health facilities. Additionally, stakeholders should send SMS reminder messages frequently at a convenient time for the recipients.

In this study, linguistic diversity was defined as the presence of different languages in the country and was declared a barrier to SMS-based appointment reminders.[17] This finding is supported by a report stating that patients' inability to communicate in their national language is a challenge for the implementation of mobile phone reminders.[34] This might be because the presence of different languages in the country makes SMS-based appointment reminder interventions and their design difficult to target all the patients or mothers. Some languages might be used only in oral communication, and the presence of cultural and linguistic differences among study subjects is a problem for SMS-based interventions.[17] So, using meaningful picture messages might be the best solution.

The lack of mobile phone ownership was mentioned as a barrier to SMS-based appointment reminders.[24] This finding is similar to reports that suggest the present lack of mobile phone ownership is a challenge for the implementation of Mhealth.[27 34] Although more than half of the world's population has access to and owns a mobile phone, a large proportion of the population shares a mobile with other people, and husbands frequently have access to the phone.[35] This is why it is difficult to deliver Mhealth interventions involving sensitive information and why husbands restrict their wives from using mobile phones. Low ownership of a mobile phone might lead to multiple users of one phone.

Strengths and limitations of the study

This study provides compiled evidence on barriers to SMS-based appointment reminders. This systematic review did not limit itself to specific groups of populations or studies that looked at specific disease patterns. At the healthcare institution level, there is a need to summarise information on challenges to mobile-based health interventions. Therefore, this study would provide evidence for stakeholders to combat the identified barriers.

This systematic review focused on varieties of interventions that were dissimilar from each other in terms of study design and population. This was a challenging task. Future systematic reviews should better focus on specific interventions and populations. Even if the impacts of SMS or mobile technology for health interventions require a search strategy, we presented it as a feature of the included studies. Due to a lack of human capital to translate, studies written in non-English languages were excluded, and meta-analysis was not conducted due to the dissimilarity of the included studies. Plus, this systematic review was limited by the studies' publication dates.

CONCLUSION

In this systematic review, we identified barriers that may potentially impede patients, mothers or other parental figures of children from receiving SMS-based appointment reminders across nine included and reviewed studies. Illiteracy, confidentiality issues, inadequate ICT infrastructure, unfamiliarity with text messaging, being a rural resident, loss of mobile phones, low mobile phone ownership, linguistic differences, work and time constraints were barriers to receiving SMS-based appointment reminders.

SMS-based text messages are an efficient and effective tool for reminding patients of medical appointments and improving health interventions. However, it is not interactive. Therefore, we recommend health policy planners work cooperatively with vendors to advance the SMS function to be interactive (two-way communication). We believed that the evidence presented in the present systematic review would provide credible evidence for better routine health interventions through mobile technology and technology adoption. This would be crucial for health policymakers, and stakeholders to address the existing identified barriers. This study would provide evidence for future research on similar topics.

Acknowledgements We want to express our gratitude to all of the studies' authors included in this systematic review.

Contributors AWD made substantial contributions to the work reported in the drafting, conception and study design. AWD and MDT made significant contributions to article screening and quality assessment. ADW had a role in article screening and revision of the manuscript, and SMW was involved in data extraction and acquisition and revising the manuscript. All the authors approved the final manuscript to be submitted for publication.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Addisalem Workie Demsash http://orcid.org/0000-0002-9356-8126
Masresha Derese Tegegne http://orcid.org/0000-0002-3156-3936
Sisay Maru Wubante http://orcid.org/0000-0001-9442-5074

REFERENCES

1 Deyo RA, Inui TS. Dropouts and broken appointments. A literature review and agenda for future research. *Med Care* 1980;18:1146–57.
2 Parikh A, Gupta K, Wilson AC, *et al*. The effectiveness of outpatient appointment reminder systems in reducing no-show rates. *Am J Med* 2010;123:542–8.
3 Hwang AS, Atlas SJ, Cronin P, *et al*. Appointment "no-shows" are an independent predictor of subsequent quality of care and resource utilization outcomes. *J Gen Intern Med* 2015;30:1426–33.
4 Berg MB, Safren SA, Mimiaga MJ, *et al*. Nonadherence to medical appointments is associated with increased plasma HIV RNA and decreased CD4 cell counts in a community-based HIV primary care clinic. *AIDS Care* 2005;17:902–7.
5 Emmink B, Frost L, Jenkins LS. Improving access to health care in a rural regional hospital in South Africa: why do patients miss their appointments? *African Journal of Primary Health Care and Family Medicine* 2017;9:1–5.
6 Cutrona SL, Choudhry NK, Fischer MA, *et al*. Targeting cardiovascular medication adherence interventions. *J Am Pharm Assoc* 2012;52:381–97.
7 Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, *et al*. Mobile phone messaging reminders for attendance at healthcare appointments. *Cochrane Database Syst Rev* 2013:CD007458.
8 Norris JB, Kumar C, Chand S, *et al*. An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. *Decis Support Syst* 2014;57:428–43.
9 Celik S, Cosansu G, Erdogan S, *et al*. Using mobile phone text messages to improve insulin injection technique and glycaemic control in patients with diabetes mellitus: a multi-centre study in turkey. *J Clin Nurs* 2015;24:1525–33.
10 Finkelstein SR, Liu N, Jani B, *et al*. Appointment reminder systems and patient preferences: patient technology usage and familiarity with other service providers as predictive variables. *Health Informatics J* 2013;19:79–90.
11 Greenhalgh T, Wherton J, Papoutsi C, *et al*. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res* 2017;19:e8775.
12 Garg SK, Lyles CR, Ackerman S, *et al*. Qualitative analysis of programmatic initiatives to text patients with mobile devices in resource-limited health systems. *BMC Med Inform Decis Mak* 2015;16:1–12.
13 McLean SM, Booth A, Gee M, *et al*. Appointment reminder systems are effective but not optimal: results of a systematic review and evidence synthesis employing realist principles. *Patient Prefer Adherence* 2016;10:479.
14 Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. *J Health Commun* 2012;17 Suppl 1:82–104.
15 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1–9.
16 Critical appraisal tools, 2021. Available: jbi.global/critical-appraisal-tools
17 Geldof M, Thiombiano BA, Wagner N. "When I receive the message, it is a sign of love": symbolic connotations of SMS messages for people living with HIV in Burkina Faso. *AIDS Care* 2021;33:810–7.
18 Nhavoto JA, Grönlund Åke, Klein GO. Mobile health treatment support intervention for HIV and tuberculosis in Mozambique: perspectives of patients and healthcare workers. *PLoS One* 2017;12:e0176051.
19 Salihu DK, Adenuga OO, Wade PD. The effect of a reminder short message service on the uptake of glaucoma screening by first-degree relatives of glaucoma patients: a randomized controlled trial. *Middle East Afr J Ophthalmol* 2019;26:196.
20 Adewuya AO, Momodu O, Olibamoyo O, *et al*. The effectiveness and acceptability of mobile telephone adherence support for management of depression in the mental health in primary care (MeHPriC) project, Lagos, Nigeria: a pilot cluster randomised controlled trial. *J Affect Disord* 2019;253:118–25.
21 Dissieka R, Soohoo M, Janmohamed A, *et al*. Providing mothers with mobile phone message reminders increases childhood immunization and vitamin A supplementation coverage in Côte d'Ivoire: a randomized controlled trial. *J Public Health Afr* 2019;10:1032.
22 Endebu T, Deksisa A, Dugasa W, *et al*. Acceptability and feasibility of short message service to improve art medication adherence among people living with HIV/AIDS receiving antiretroviral treatment at Adama hospital medical College, central Ethiopia. *BMC Public Health* 2019;19:1–11.
23 Schroeder K, Maiarana J, Gisiri M, *et al*. Caregiver acceptability of mobile phone use for pediatric cancer care in Tanzania: cross-sectional questionnaire study. *JMIR Pediatr Parent* 2021;4:e27988.
24 Mekonnen ZA, Gelaye KA, Were MC, *et al*. Acceptability, barriers and facilitators of mobile text message reminder system implementation in improving child vaccination: a qualitative study in Northwest Ethiopia. *J Multidiscip Healthc* 2021;14:605–16.
25 Moodley J, Constant D, Botha MH, *et al*. Exploring the feasibility of using mobile phones to improve the management of clients with cervical cancer precursor lesions. *BMC Womens Health* 2019;19:1–10.
26 Katz M, Fitzek FH. *WiMAX evolution: emerging technologies and applications*. John Wiley & Sons, 2009.
27 Muller N, McMahon SA, De Neve J-W, *et al*. Facilitators and barriers to the implementation of a mobile health Wallet for pregnancy-related health care: a qualitative study of stakeholders' perceptions in Madagascar. *PLoS One* 2020;15:e0228017.
28 Chowdhury KP. Literacy and primary education: world bank. *Human Resources Development and Operations Policy* 1995.
29 Househ M. The role of short messaging service in supporting the delivery of healthcare: an umbrella systematic review. *Health Informatics J* 2016;22:140–50.
30 Gurupur VP, Wan TTH. Challenges in implementing mHealth interventions: a technical perspective. *Mhealth* 2017;3:32.
31 Rothstein JD, Jennings L, Moorthy A, *et al*. Qualitative assessment of the feasibility, usability, and acceptability of a mobile client data APP for community-based maternal, neonatal, and child care in rural Ghana. *Int J Telemed Appl* 2016;2016:1–14.
32 Ngabo F, Nguimfack J, Nwaigwe F, *et al*. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. *Pan Afr Med J* 2012;13:31.
33 Haberer JE, Kiwanuka J, Nansera D, *et al*. Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. *AIDS Behav* 2010;14:1294–301.
34 Bigna JJR, Noubiap JJN, Plottel CS, *et al*. Barriers to the implementation of mobile phone reminders in pediatric HIV care: a pre-trial analysis of the Cameroonian more care study. *BMC Health Serv Res* 2014;14:1–7.
35 Silver L, Huang C, Taylor K. *In emerging economies, smartphone and social media users have broader social networks*. Pew Research Center, 2019.

BMJ Health & Care Informatics

Extension of the Unified Theory of Acceptance and Use of Technology 2 model for predicting mHealth acceptance using diabetes as an example: a cross-sectional validation study

Patrik Schretzlmaier,[1] Achim Hecker,[2,3] Elske Ammenwerth [1]

To cite: Schretzlmaier P, Hecker A, Ammenwerth E. Extension of the Unified Theory of Acceptance and Use of Technology 2 model for predicting mHealth acceptance using diabetes as an example: a cross-sectional validation study. *BMJ Health Care Inform* 2022;**29**:e100640. doi:10.1136/bmjhci-2022-100640

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100640).

Received 23 July 2022
Accepted 28 October 2022

[1]Institute of Medical Informatics, UMIT TIROL—Private University for Health Sciences and Health Technology, Hall, Tirol, Austria
[2]Institute for Management and Economics in Healthcare, UMIT TIROL—Private University for Health Sciences and Health Technology, Hall, Tirol, Austria
[3]DBU Digital Business University of Applied Sciences, Berlin, Germany

Correspondence to
Dr Elske Ammenwerth; elske.ammenwerth@umit-tirol.at

ABSTRACT

Objectives Mobile health applications are instrumental in the self-management of chronic diseases like diabetes. Technology acceptance models such as Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) have proven essential for predicting the acceptance of information technology. However, earlier research has found that the constructs "perceived disease threat" and "trust" should be added to UTAUT2 in the mHealth acceptance context. This study aims to evaluate the extended UTAUT2 model for predicting mHealth acceptance, represented by behavioral intention, using mobile diabetes applications as an example.

Methods We extended UTAUT2 with the additional constructs "perceived disease threat" and "trust". We conducted a web-based survey in German-speaking countries focusing on patients with diabetes and their relatives who have been using mobile diabetes applications for at least 3 months. We analysed 413 completed questionnaires by structural equation modelling.

Results We could confirm that the newly added constructs "perceived disease threat" and "trust" indeed predict behavioural intention to use mobile diabetes applications. We could also confirm the UTAUT2 constructs "performance expectancy" and "habit" to predict behavioural intention to use mobile diabetes applications. The results show that the extended UTAUT2 model could explain 35.0% of the variance in behavioural intention.

Discussion Even if UTAUT2 is well established in the information technologies sector to predict technology acceptance, our results reveal that the original UTAUT2 should be extended by "perceived disease threat" and "trust" to better predict mHealth acceptance.

Conclusion Despite the newly added constructs, UTAUT2 can only partially predict mHealth acceptance. Future research should investigate additional mHealth acceptance factors, including how patients perceive trust in mHealth applications.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ mHealth applications are essential for comprehensive self-management of chronic diseases.
⇒ The use of mHealth applications depends significantly on their acceptance, which can be predicted using technology acceptance models.
⇒ The UTAUT2 model has proven suitable for predicting technology acceptance in various information technology domains.

WHAT THIS STUDY ADDS

⇒ "Perceived disease threat" and "trust" are relevant in predicting the acceptance of mobile diabetes applications and should be added to Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) when used in this context.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ "Performance expectancy", "habit", "perceived disease threat" and "trust" are relevant for accepting mobile diabetes applications. Therefore, these factors should be considered when developing new mHealth applications in this context.
⇒ External conditions, such as country-specific financial support for mHealth users from, for example, statutory health insurances, which helps to fund the required mHealth applications, should be addressed in future mHealth acceptance studies.

INTRODUCTION

mHealth acceptance

Mobile health (mHealth) applications, especially the so-called lifestyle apps such as fitness apps, have become increasingly popular, specifically among younger people, due to growing health awareness.[1 2] Besides lifestyle apps, mHealth applications such as continuous glucose monitoring systems (CGMs) are instrumental for the self-management of chronic diseases like diabetes mellitus.[3 4] Different studies have shown that using mHealth applications leads to improved self-management and better health among people with chronic diseases.[5 6] This is especially true in the case of diabetes, which is one of the most frequently occurring chronic diseases worldwide.[7] mHealth applications

for patients with diabetes can support sustained self-management and help maintain lower long-term glucose levels.[3 6 8]

Despite the benefits associated with mHealth applications, there are several reasons why they are not used, such as difficulties in their control[9] or acceptance problems.[10–12] User acceptance can be described as 'the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support'.[13] Several studies have shown that, especially for people with type 2 diabetes, the acceptance of mHealth self-management applications is noticeably low.[8 14]

Theoretical background

Technology acceptance models such as the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) have been developed to predict the acceptance of information technologies in health informatics and other fields of application.[15–17]

The UTAUT2 model was established in 2012 for use in a consumer context.[18] In contrast to the previous technology acceptance models, UTAUT2 used additional exogenous constructs "habit", "hedonic motivation" and "price value" to predict the endogenous construct "behavioral intention", which is understood as an expression of technology acceptance.[18] With the focus on the individuals and their needs, UTAUT2 is particularly suitable for predicting the acceptance of mHealth applications such as mobile diabetes applications.[18 19] However, it is still not as widely used in mHealth acceptance studies as other technology acceptance models. Some studies using UTAUT2 have pointed out that essential aspects such as health-related factors[19 20] or factors related to trust in the data collected[2 21] are missing. In a previous qualitative study, we could confirm the general suitability of the UTAUT2 model in the field of mHealth self-management applications but identified some missing aspects, such as the awareness of the perceived threat of disease and credibility in the data collected by the application for predicting mHealth acceptance.[22] Therefore, we proposed adding the following constructs to the UTAUT2 model: "perceived disease threat" and "trust". The construct "trust" is associated with the belief that people accept uncertainties due to positive expectations.[2] It is used to determine the data credibility and trustworthiness of the mobile health application, which is particularly important for behavioural intention and long-term use of mHealth applications.[2 22]

Furthermore, when patients face health-threatening situations, they are more open to new health technologies.[11 12] Especially with chronic diseases like diabetes, the individual awareness of the risk and limitations for their health, reflected by the construct "perceived disease threat", is a significant driver for acceptance and use of mHealth applications.[20 22] Few studies have used the UTAUT2 model to predict mHealth acceptance to date, and these studies have not yet considered the two constructs of "perceived disease threat" and "trust".[19 22]

Study objectives

This study aims to validate whether the exogenous UTAUT2 constructs, combined with the additional constructs "perceived disease threat" and "trust", can predict mHealth acceptance using mobile diabetes applications as an example.

Hypotheses development and proposed research model

Figure 1 shows the proposed research model using the exogenous UTAUT2 constructs and additional constructs "perceived disease threat" (PDT) and "trust" (TR) for predicting the endogenous construct "behavioral intention" (BI) to use mHealth applications. Although this

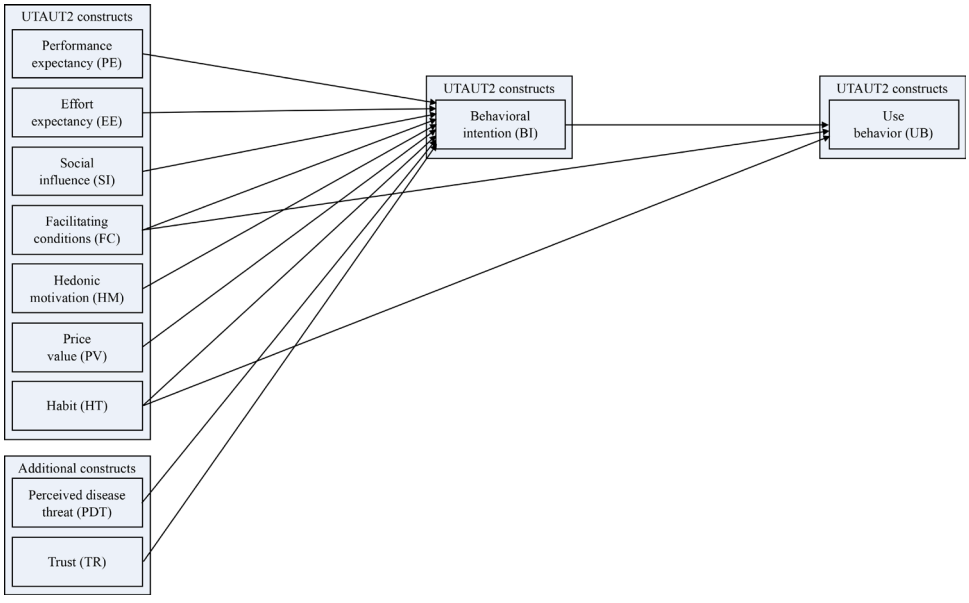


Figure 1 Research model based on Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) and additional constructs for predicting mHealth acceptance.[18]

study focused on the acceptance (BI) of mobile diabetes applications, we also included the endogenous construct "use behavior" (UB) in the analysis to validate the exogenous constructs in the complete UTAUT2 model.

Based on the existing UTAUT2 model, we adopted the relationships between the exogenous constructs "performance expectancy" (PE), "effort expectancy" (EE), "social influence" (SI), "facilitating conditions" (FC), "hedonic motivation" (HM), "price value" (PV) and "habit" (HT) and the endogenous construct "behavioral intention" (BI).[18] In addition, the factors PDT and TR have been shown in various studies to predict the acceptance of mHealth applications.[1 2 11 12 20 22] This leads to the following hypothesis: PE, EE, SI, FC, HM, PV, HT, PDT and TR affect the BI to use mobile diabetes applications.

METHODS

Study design

To validate the proposed extension of the UTAUT2 model, we adopted a cross-sectional study design based on data from active mHealth users collected in an online survey.

Questionnaire

We used a web-based questionnaire based on a previously validated German translation of the UTAUT2 questionnaire.[23] We slightly adapted the wording of the items to match the area of mobile diabetes applications.

We extended the questionnaire with already validated English items for the constructs PDT[24] and TR.[25] We used back-to-back translation by two independent translators fluent in English and German to translate these items, as no German translation was available.

The central part of the questionnaire consisted of the validated UTAUT2 items combined with the validated items from the constructs PDT and TR to predict mHealth acceptance (see table 1).

We used a seven-point Likert scale ranging from 1 'strongly disagree' to 7 'strongly agree' to measure the items of the reflective measurement models.

In addition, we assessed information such as users' experience with mobile diabetes applications, type of diabetes, use behaviour and sociodemographic data using single-item measurement. Overall, the questionnaire comprised 42 items.

We conducted a qualitative pretest to confirm content validity and understandability of the translated questionnaire. We discussed the questionnaire with five academic experts from the field of quantitative research and five users of mobile diabetes applications. We used the think-aloud method and captured all feedback. The results from this qualitative pretest were collected and discussed with all coauthors to reach a shared consensus before implementation. We slightly revised the questionnaire, for example, by adding some examples.

Participants

We used convenience sampling to recruit mobile diabetes application users from Austria and Germany. We included persons with diabetes type 1, type 2, and others or persons caring for relatives (eg, child) with diabetes who were 18 years or older and used a mobile diabetes application (eg, CGMs) for at least 3 months. We only included active mobile diabetes application users, as some constructs, such as HT, require current use.[2 18] We primarily used social media to recruit participants for our web-based questionnaire. In addition, we teamed up with gatekeepers in diabetes associations, support groups and direct contacts to medical staff in diabetes outpatient clinics to encourage patients to participate in the online survey.

Participation in the online survey was voluntary, anonymous and could be discontinued anytime. We used browser session ID blocking and cookie settings to prevent multiple participation from the same individuals.

The inverse square root method was used to calculate the required sample size.[26] Using a power of 80%, significance level $p<0.05$, and a minimum path coefficient of $p_{min}=0.185$ based on studies with similar complexity,[2] the calculated minimum sample size was 181 participants.

Data analysis

We used partial least squares structural equation modelling (PLS-SEM) for data analysis because this method is particularly suitable for exploratory research with high model complexity and has already been used and established in similar studies.[1 2 27] We used SmartPLS3 software for structural equation modelling.

We followed the data analysis approach described by Hair *et al*,[28] divided into measurement and structural model evaluation. After completing the measurement and structural model evaluation, we conducted an additional moderator analysis using the UTAUT2 moderators 'age', 'gender' and 'experience' to evaluate any potential moderator effect.

RESULTS

Descriptive data

Overall, 514 persons participated in the web-based survey, of which only 413 gave their consent, completed the questionnaire, and met the inclusion criteria. The sample demographic characteristics are shown in table 2.

Measurement model evaluation

We started evaluating the reflective measurement model's convergent validity by checking the indicators' loadings to assess indicator reliability. However, the loadings of the items FC4, HT2, PDT1 and PV1 did not satisfy the required value of 0.708.[27] Based on Hair *et al*,[28] items with loadings below 0.40 should be eliminated. Therefore, we deleted PDT1 with a loading of 0.156 from the measurement model for the following investigations. In addition, items with loadings between 0.40 and 0.70 should only be deleted if this will increase composite reliability, which was not the case for FC4, HT2 and PV1.[28] We assessed internal consistency reliability by checking composite

Table 1 Constructs and items used in the web-based questionnaire

Construct	Items	Cronbach's alpha	Source adapted from
Performance expectancy (PE)	PE1: I find mobile diabetes applications useful in my daily life. PE2: Using mobile diabetes applications increases my chances of achieving things that are important to me. PE3: Using mobile diabetes applications helps me accomplish things more quickly. PE4: Using mobile diabetes applications increases my productivity.	0.951*	Venkatesh *et al*.[18], Harborth and Pape[23]
Effort expectancy (EE)	EE1: Learning how to use mobile diabetes applications is easy for me. EE2: My interaction with mobile diabetes applications is clear and understandable. EE3: I find mobile diabetes applications easy to use. EE4: It is easy for me to become skillful at using mobile diabetes applications.	0.922*	Venkatesh *et al*.[18], Harborth and Pape[23]
Social influence (SI)	SI1: People who are important to me think that I should use mobile diabetes applications. SI2: People who influence my behavior think that I should use mobile diabetes applications. SI3: People whose opinions that I value prefer that I use mobile diabetes applications.	0.948*	Venkatesh *et al*.[18], Harborth and Pape[23]
Facilitating conditions (FC)	FC1: I have the resources necessary to use mobile diabetes applications. FC2: I have the knowledge necessary to use mobile diabetes applications. FC3: Mobile diabetes applications are compatible with other technologies I use. FC4: I can get help from others when I have difficulties using mobile diabetes applications.	0.733*	Venkatesh *et al*.[18], Harborth and Pape[23]
Hedonic motivation (HM)	HM1: Using mobile diabetes applications is fun. HM2: Using mobile diabetes applications is enjoyable. HM3: Using mobile diabetes applications is very entertaining.	0.937*	Venkatesh *et al*.[18], Harborth and Pape[23]
Price value (PV)	PV1: Mobile diabetes applications are reasonably priced. PV2: Mobile diabetes applications are a good value for the money. PV3: At the current price, mobile diabetes applications provide a good value.	0.867*	Venkatesh *et al*.[18], Harborth and Pape[23]
Habit (HT)	HT1: The use of mobile diabetes applications has become a habit for me. HT2: I am addicted to using mobile diabetes applications. HT3: I must use mobile diabetes applications.	0.879*	Venkatesh *et al*.[18], Harborth and Pape[23]
Behavioral intention (BI)	BI1: I intend to continue using mobile diabetes applications in the future. BI2: I will always try to use mobile diabetes applications in my daily life. BI3: I plan to continue to use mobile diabetes applications frequently.	0.898*	Venkatesh *et al*.[18], Harborth and Pape[23]
Use behavior (UB)	Please choose your usage frequency for mobile diabetes applications: Never Once a month Several times a month Once a week Several times a week Once a day Several times a day Once an hour Several times an hour All the time	1.000*	Venkatesh *et al*.[18], Harborth and Pape[23]

Continued

Table 1 Continued

Construct	Items	Cronbach's alpha	Source adapted from
Perceived disease threat (PDT)†	PDT1: I am aware that my blood sugar control is not optimal. PDT2: I am very concerned about my blood sugar. PDT3: I am very concerned about diabetes-associated complications.	0.743‡	Zhang *et al*.[24]
Trust (TR)†	TR1: I trust my mobile diabetes application. TR2: I find mobile diabetes applications reliable in conducting health services. TR3: I feel that mobile diabetes applications are safe for receiving reliable medical information. TR4: I trust mobile diabetes applications' commitment to satisfy my medical information needs.	0.869§	Lee *et al*.[25]

*Cronbach's alpha retrieved from Harborth and Pape.[23]
†New constructs added to UTAUT2 for this study.
‡Cronbach's alpha retrieved from Zhang *et al*.[24]
§Cronbach's alpha retrieved from Lee *et al*.[25]
UTAUT2, Unified Theory of Acceptance and Use of Technology 2.

reliability and Cronbach's alpha in the second step, both in the recommended range for exploratory research, as shown in table 3. In the third step, we evaluated convergent validity by assessing the average variance extracted (AVE). The values for AVE were above the required 0.50.[27] In the last step, we assessed discriminant validity by checking the heterotrait–monotrait (HTMT) ratio of the correlations for which all values were in the interval (0.028, 0.747), satisfying the HTMT requirements to be below 0.90[27] (see online supplemental material 1). Thus, the evaluation of the measurement model proved that our data satisfied the requirements for reliability and validity. However, UB was excluded from the measurement model evaluation since the test criteria do not apply to single-item constructs.[28]

Structural model evaluation

In the first step of structural model evaluation, we assessed collinearity to ensure no bias in the regression results using the variance inflation factor (VIF). The values for VIF ideally should be below 3,[27] which could be verified by our data with values in the interval (1.061, 1.741) (see online supplemental material 2). We assessed the path coefficients by checking their p values using the bootstrapping method (see figure 2). We applied a significance level of $p<0.05$, where we could identify that only PDT (0.091, p=0.019, f^2=0.012), TR (0.145, p=0.034, f^2=0.022), PE (0.285, $p<0.001$, f^2=0.076) and HT (0.171, p=0.025, f^2=0.027), had a significant impact on BI (see online supplemental material 3). In addition, we could identify a significant impact of HT (0.362, $p<0.001$, f^2=0.113) and FC (−0.110, p=0.049, f^2=0.012) on UB (see online supplemental material 4). Next, we assessed the adjusted coefficient of determination ($R^2_{adj.}$), which measures the models' predictive power and explains the variance of the endogenous construct.[27] Our extended UTAUT2 model could explain 35.0% of the variance in BI and 14.4% in UB (see figure 2) compared with the original UTAUT2 model without the constructs PDT and TR, which explained 32.8% of the variance in BI and 14.4% in UB (see online supplemental material 5). In addition, we assessed the predictive relevance (Q^2) of the extended UTAUT2 model for BI and UB using blindfolding procedures. Our model achieved a Q^2=0.274 for BI, implying medium predictive accuracy, and a Q^2=0.129 for UB, implying low predictive accuracy.[27] Thus, our results indicate that the extended UTAUT2 model is suitable for predicting the acceptance of mobile diabetes applications. The additional moderator analysis revealed no significant effect of the UTAUT2 moderators on the relationships between endogenous and exogenous constructs (see online supplemental material 6).

DISCUSSION

Principal findings and comparison with prior work

As a starting point for such an extension, we could confirm the relevance of the newly added constructs PDT and TR for predicting mHealth acceptance. Both constructs have already been individually highlighted as relevant in earlier mHealth acceptance studies.[1 2 20] In particular, TR has been emphasised as a significant driver for accepting mHealth applications as they need to be reliable and trustworthy.[2 12 21] Several studies have shown that TR is crucial for BI, especially for applications that impact personal health, such as mobile diabetes applications.[1 2 12] PDT has also been identified as significant to the acceptance of mHealth applications, particularly for chronic diseases.[12 19 24] Various studies have emphasised that people who are aware and concerned about their poor health conditions are more open to new technologies.[20 24 29] To summarise, PDT and TR need to be considered for a UTAUT2 extension to predict mHealth acceptance better.

Table 2 Sample demographic characteristics

Characteristics	Mobile diabetes app users (n=413)
Gender, n (%)	
Female	256 (62.0%)
Male	152 (36.8%)
Diverse	0 (0.0%)
Not mentioned	5 (1.2%)
Age, n (%)	
18–24 years	15 (3.6%)
25–34 years	68 (16.5%)
35–44 years	102 (24.7%)
45–54 years	108 (26.2%)
55–64 years	84 (20.3%)
65 years and older	32 (7.7%)
Not mentioned	4 (1.0%)
Type of diabetes, n (%)	
Type 1	344 (83.3%)
Type 2	64 (15.5%)
Others	5 (1.2%)
Disease duration, years	
Mean (SD)	18.17 (SD 14.77)
Range (median)	0.25–63.00 (15.00)
Duration of use, months	
Mean (SD)	31.30 (SD 26.18)
Range (median)	3.00–240.00 (24.00)
Type of mobile diabetes app, n (%)	
Smartphone app (eg, diabetes diary)	199 (48.18%)
Continuous glucose measurement system	383 (92.7%)
Smart insulin pump (eg, closed-loop system)	79 (19.1%)
Others	0 (0.0%)
Frequency of use, n (%)	
Never	0 (0.0%)
Once a month	1 (0.2%)
Several times a month	3 (0.7%)
Once a week	2 (0.5%)
Several times a week	6 (1.5%)
Once a day	3 (0.7%)
Several times a day	154 (37.3%)
Once an hour	25 (6.1%)
Several times an hour	59 (14.3%)
All the time	160 (38.7%)

We also identified a significant impact of the UTAUT2 constructs, PE and HT, in predicting mHealth acceptance, where PE showed the strongest effect, which is in line with several mHealth acceptance studies.[20 30]

In contrast to the previous qualitative study,[22] we could not confirm the significance of the remaining exogenous UTAUT2 constructs EE, SI, FC, HM and PV for predicting mHealth acceptance. These findings are consistent with other mHealth acceptance studies using UTAUT2 that also showed no significant impact of those constructs on BI.[12 20 30]

There could be different reasons why proven UTAUT2 constructs are not showing significance in this study. First, the motivation of people with chronic diseases to use mobile applications is probably different from those of the average technology consumer. While aspects such as fun (HM), convenience (EE), low price (PV), social influence (SI) and support (FC) are more relevant in the consumer context, the focus for mHealth applications is on the actual health benefit (PE) and the possibility of integrating the applications into everyday life (HT).[11 12 18 30] Second, the relevance of these aspects might depend on the mHealth application and how it is provided to the patient. For example, according to the 'National Association of Health Insurers' (German: GKV-Spitzenverband) in Germany and the 'Main Association of Austrian Social Insurance Institutions' (German: Hauptverband der oesterreichischen Sozialversicherungstraeger), CGM systems are considered medical aids that physicians usually prescribe, and whose costs are covered by statutory health insurances.[22] Thus, the application used is mainly specified by the statutory health insurance, which makes factors such as the cost (PV) and personal recommendations (SI) less relevant for patients. As 92.7% of participants used CGMs, this could explain the lack of influence of several UTAUT2 constructs in this study. Additionally, the German Federal Institute for Drugs and Medical Devices (BfArM) has released the so-called digital health applications (DiGA) registry, containing digital medical products (web applications, native apps) which have undergone an appropriate evaluation procedure and can be prescribed by physicians and psychotherapists at the expense of the statutory health insurances.[20] Third, from the perspective of several authors of mHealth acceptance studies using UTAUT2, constructs such as PV and HM are not applicable.[20 21 30] In summary, all this may explain why not all UTAUT2 constructs are relevant in the context of mHealth acceptance. Thus, this points to the need for a larger UTAUT2 extension for mHealth acceptance.

Using our extended UTAUT2 model, we increased the explained variance in BI from 32.8% in the original UTAUT2 model to 35.0%. This aligns with other mHealth acceptance studies extending UTAUT2, where the explained variance in BI ranged from 19.4%[2] to 56.0%.[20] Thus, the explained variance in BI of $R^2 = 0.350$ is comparable to results from other studies and is within the medium range.

Finally, the additional moderator analysis revealed no significant moderating effect, which aligns with other mHealth acceptance studies not showing a moderating effect.[1 29]

Table 3 Evaluation of the measurement model

		Convergent validity			Internal consistency reliability		Discriminant validity
		Loadings	Indicator reliability	AVE	Composite reliability	Cronbach's alpha	
Construct	Item	> 0.708	> 0.50	> 0.50	> 0.70	0.60–0.95	HTMT ratio below 0.90
BI	BI1	0.872	0.760	0.835	0.938	0.901	Yes
	BI2	0.938	0.880				
	BI3	0.929	0.863				
EE	EE1	0.899	0.808	0.773	0.931	0.902	Yes
	EE2	0.909	0.826				
	EE3	0.901	0.812				
	EE4	0.804	0.646				
FC	FC1	0.776	0.602	0.525	0.812	0.695	Yes
	FC2	0.757	0.573				
	FC3	0.809	0.654				
	FC4	0.521	0.271				
HM	HM1	0.891	0.794	0.705	0.876	0.807	Yes
	HM2	0.899	0.808				
	HM3	0.716	0.513				
HT	HT1	0.798	0.637	0.561	0.792	0.628	Yes
	HT2	0.672	0.452				
	HT3	0.771	0.594				
PDT	PDT2	0.876	0.767	0.844	0.915	0.817	Yes
	PDT3	0.936	0.876				
PE	PE1	0.789	0.623	0.666	0.888	0.833	Yes
	PE2	0.801	0.642				
	PE3	0.835	0.697				
	PE4	0.838	0.702				
PV	PV1	0.663	0.440	0.706	0.876	0.789	Yes
	PV2	0.910	0.828				
	PV3	0.922	0.850				
SI	SI1	0.875	0.766	0.768	0.909	0.850	Yes
	SI2	0.887	0.787				
	SI3	0.867	0.752				
TR	TR1	0.788	0.621	0.662	0.887	0.830	Yes
	TR2	0.784	0.615				
	TR3	0.817	0.667				
	TR4	0.864	0.746				

BI, behavioural intention; EE, effort expectancy; FC, facilitating conditions; HM, hedonic motivation; HT, habit; PDT, perceived disease threat; PE, performance expectancy; PV, price value; SI, social influence; TR, trust.

Strengths and limitations

With our quantitative study design, we validated the proposed extension of the UTAUT2 model for predicting mHealth acceptance of mobile diabetes applications. Using validated scales combined with a qualitative pretest, we could ensure the quality of our web-based questionnaire.

The focus of this study was to validate the constructs relevant for predicting mHealth acceptance (BI) in the UTAUT2 model. We thus only briefly touched on the construct of UB. UB is subject to some weaknesses due to its operationalisation as an ordinal-scaled single item in the validated German translation of the UTAUT2 questionnaire. In addition, several UTAUT and UTAUT2

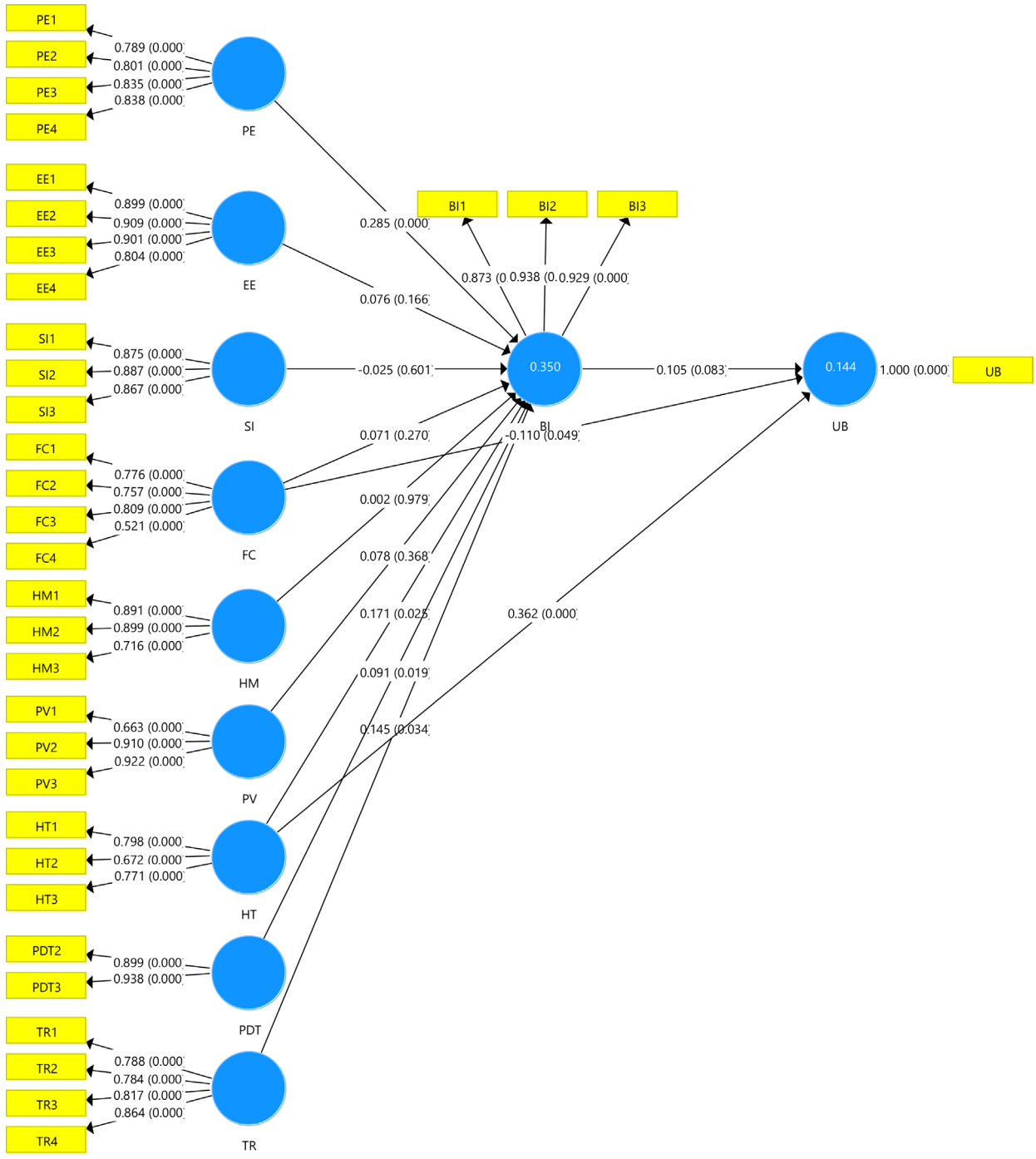


Figure 2 Extended Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model, including path coefficients and adjusted coefficient of determination.

studies focusing on mHealth acceptance have operationalised BI only and have not considered UB.[2 20 31] However, in mHealth acceptance studies that operationalised UB, the values for the explained variance (R^2) in UB were also relatively low, for example, $R^2 = 0.111$ in Dou *et al*[29] or $R^2 = 0.320$ in Fitrianie *et al*,[21] which is consistent with our results of $R^2 = 0.144$. We thus concentrated on BI in our statistical analysis.

Using a convenience sample-based recruitment process via social media and directly through, for example, diabetes outpatient clinics, we recruited a diverse sample of different ages, genders, diabetes types and experiences. Using these different recruitment channels, we reduced the risk of sampling bias, thus ensuring that we reached all relevant user groups. However, since all participants used the same link to access the web-based questionnaire (regardless of whether they accessed it via social media or another recruitment channel), it was impossible to distinguish which proportion of the sample derived from social media platforms and which, for example, from diabetes outpatient clinics.

With 413 participants, our study achieved sufficient power to perform PLS-SEM and predict mHealth acceptance. Our study only involved active and experienced mobile diabetes application users. Thus, we did not include non-users or seldom users, making our result generalisable to real mHealth users who use the mHealth application as part of their chronic disease self-management, have similar levels of experience with mHealth applications, and come from countries with comparable healthcare systems, technical infrastructure, cultural and socioeconomic background like the convenience sample used. In particular, this applies to mHealth self-management applications, such as diet or exercise apps designed for various chronic diseases, for example, type 2 diabetes, hypertension, obesity and others.[32]

Implications for future research

Our results showed the relevance of PE, HT, PDT and TR for accepting mobile diabetes applications. Therefore, mHealth providers should consider addressing and implementing these factors when developing new mHealth applications in this context.

Although we could verify two new relevant constructs, our extended UTAUT2 model still explained only 35.0% of the variance in BI. Further research thus needs to identify additional factors for predicting mHealth acceptance that have not yet been identified and considered in the extended UTAUT2 model.

In addition, future mHealth acceptance research in chronic diseases should consider external conditions, such as country-specific financial support for mHealth users from, for example, statutory health insurances, which helps to fund the required mHealth applications.

Future research may also investigate if considering additional dimensions to our extended UTAUT2 model, for example, from the diffusion of innovations theory, could contribute to the explained variance in BI.

Our study focused on active and experienced mobile diabetes application users. Future research may also involve patients with diabetes who do not use mobile diabetes applications. This could help understanding barriers and facilitators to accept and use mobile diabetes applications and identify potential additional predictors.

Since other studies have identified previous use as a strong predictor of future behaviour,[20] this predictor should be considered in future mHealth acceptance studies using UTAUT2.

CONCLUSIONS

We could confirm that the additional constructs PDT and TR need to be added to UTAUT2 to predict mHealth acceptance, especially for chronic diseases such as diabetes. However, our results also show that not all constructs of the UTAUT2 model can predict mHealth acceptance.

Despite the newly added constructs, UTAUT2 can only partially predict mHealth acceptance. Future research should investigate additional mHealth acceptance factors, including how patients perceive trust in mHealth applications.

Acknowledgements We gratefully acknowledge the Austrian Diabetes Association, Diabetics Lower Saxony e.V., Tirol Clinics GmbH, and Dr. Stefan Richter for their help with mobile diabetes app user recruitment. In addition, we would like to thank all online survey participants.

Contributors PS and EA planned the study. PS conducted the data collection and analysis and drafted and revised the paper. PS and AH planned the statistical analysis. EA and AH helped in reviewing the results of the pretests, in data analysis, and in drafting the paper. PS acts as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by research committee for scientific, ethical questions of the UMIT TIROL - Private University for Health Sciences and Health Technology (RCSEQ 2805/20). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD

Elske Ammenwerth http://orcid.org/0000-0002-2365-4766

REFERENCES

1 Akdur G, Aydin MN, Akdur G. Adoption of mobile health apps in dietetic practice: case study of diyetkolik. *JMIR Mhealth Uhealth* 2020;8:e16911.
2 Schomakers E-M, Lidynia C, Vervier LS, *et al*. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. *JMIR Mhealth Uhealth* 2022;10:e27095.
3 Hood M, Wilson R, Corsica J, *et al*. What do we know about mobile applications for diabetes self-management? A review of reviews. *J Behav Med* 2016;39:981–94.
4 Virella Pérez YI, Medlow S, Ho J, *et al*. Mobile and web-based Apps that support self-management and transition in young people with chronic illness: systematic review. *J Med Internet Res* 2019;21:e13579.
5 Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2020;8:e21759.
6 Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019;7:e12297.
7 International Diabetes Federation. *IDF diabetes atlas*. 10th edition, 2021.
8 Trawley S, Baptista S, Browne JL, *et al*. The use of mobile applications among adults with type 1 and type 2 diabetes: results

from the second Miles-Australia (Miles-2) study. *Diabetes Technol Ther* 2017;19:730–8.
9 Bentley CL, Powell L, Potter S, *et al*. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. *JMIR Mhealth Uhealth* 2020;8:e16203.
10 Scheibe M, Reichelt J, Bellmann M, *et al*. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. *Med 2 0* 2015;4:e1.
11 Breil B, Kremer L, Hennemann S, *et al*. Acceptance of mHealth apps for self-management among people with hypertension. *Stud Health Technol Inform* 2019;267:282–8.
12 Jacob C, Sezgin E, Sanchez-Vazquez A, *et al*. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. *JMIR Mhealth Uhealth* 2022;10:e36284.
13 Dillon A. User acceptance of information technology. London Taylor and Francis; 2001: 10.
14 Bults M, van Leersum CM, Olthuis TJJ, *et al*. Barriers and drivers regarding the use of mobile health apps among patients with type 2 diabetes mellitus in the Netherlands: explanatory sequential design study. *JMIR Diabetes* 2022;7:e31451.
15 Venkatesh V, Morris MG, Davis GB, *et al*. User acceptance of information technology: toward a unified view. *MIS Quarterly* 2003;27:425–78.
16 Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int J Man Mach Stud* 1993;38:475–87.
17 Ammenwerth E. Technology acceptance models in health informatics: TAM and UTAUT. *Stud Health Technol Inform* 2019;263:64–71.
18 Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Quarterly* 2012;36:157–78.
19 Alam MMD, Alam MZ, Rahman SA, *et al*. Factors influencing mHealth adoption and its impact on mental well-being during COVID-19 pandemic: a SEM-ANN approach. *J Biomed Inform* 2021;116:103722.
20 Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. *JMIR Cardio* 2022;6:e31617.
21 Fitrianie S, Horsch C, Beun RJ, *et al*. Factors affecting user's behavioral intention and use of a mobile-phone-delivered cognitive behavioral therapy for insomnia: a small-scale UTAUT analysis. *J Med Syst* 2021;45:110.
22 Schretzlmaier P, Hecker A, Ammenwerth E. Suitability of the unified theory of acceptance and use of technology 2 model for predicting mHealth acceptance using diabetes as an example: qualitative methods triangulation study. *JMIR Hum Factors* 2022;9:e34918.
23 Harborth D, Pape S. German translation of the unified theory of acceptance and use of technology 2 (UTAUT2) questionnaire. *SSRN Journal* 2018;116.
24 Zhang Y, Liu C, Luo S, *et al*. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. *J Med Internet Res* 2019;21:e15023.
25 Lee W-I, Fu H-P, Mendoza N, *et al*. Determinants impacting user behavior towards emergency use intentions of m-Health services in Taiwan. *Healthcare* 2021;9. doi:10.3390/healthcare9050535. [Epub ahead of print: 03 05 2021].
26 Kock N. Minimum Sample Size Estimation in PLS-SEM: An Application in Tourism and Hospitality Research. In: Ali F, ed. *Applying partial least squares in tourism and hospitality research*. Bingley: Emerald Publishing Limited, 2019: 1–16.
27 Hair JF, Risher JJ, Sarstedt M, *et al*. When to use and how to report the results of PLS-SEM. *EBR* 2019;31:2–24.
28 Hair JF, Hult GTM, Ringle CM. *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles: Sage Publ, 2014.
29 Dou K, Yu P, Deng N, *et al*. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. *JMIR Mhealth Uhealth* 2017;5:e177.
30 Salgado T, Tavares J, Oliveira T. Drivers of mobile health acceptance and use from the patient perspective: survey study and quantitative model development. *JMIR Mhealth Uhealth* 2020;8:e17588.
31 Apolinário-Hagen J, Menzel M, Hennemann S, *et al*. Acceptance of mobile health Apps for disease management among people with multiple sclerosis: web-based survey study. *JMIR Form Res* 2018;2:e11977.
32 Debon R, Coleone JD, Bellei EA, *et al*. Mobile health applications for chronic diseases: a systematic review of features for lifestyle improvement. *Diabetes Metab Syndr* 2019;13:2507–12.

BMJ Health & Care Informatics

Pilot trial comparing COVID-19 publication database to conventional online search methods

Camille Torfs-Leibman,[1] Takamaru Ashikaga,[2] David Krag ,[3] Shania Lunna,[3] Sarah Robtoy,[1] Rachel Bombardier[4]

To cite: Torfs-Leibman C, Ashikaga T, Krag D, *et al*. Pilot trial comparing COVID-19 publication database to conventional online search methods. *BMJ Health Care Inform* 2022;**29**:e100616. doi:10.1136/bmjhci-2022-100616

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100616).

Received 27 June 2022
Accepted 15 October 2022

[1]University of Vermont, Burlington, Vermont, USA
[2]Department of Biomedical Statistics, University of Vermont, Burlington, Vermont, USA
[3]Department of Surgery, University of Vermont, Burlington, Vermont, USA
[4]University of Vermont Larner College of Medicine, Burlington, Vermont, USA

Correspondence to
Camille Torfs-Leibman; ctorfsleibman@gmail.com

ABSTRACT

Background and objectives Literature review using search engines results in a list of manuscripts but does not provide the content contained in the manuscripts. Our goal was to evaluate user performance-based criteria of concept retrieval accuracy and efficiency using a new database system that contained information extracted from 1000 COVID-19 articles.

Methods A sample of 17 students from the University of Vermont were randomly assigned to use the COVID-19 publication database or their usual preferred search methods to research eight prompts about COVID-19. The relevance and accuracy of the evidence found for each prompt were graded. A Cox proportional hazards' model with a sandwich estimator and Kaplan-Meier plots were used to analyse these data in a time-to-correct answer context.

Results Our findings indicate that students using the new information management system answered significantly more prompts correctly and, in less time, than students using conventional research methods. Bivariate models for demographic factors indicated that previous research experience conferred an advantage in study performance, though it was found to be independent from the assigned research method.

Conclusions The results from this pilot randomised trial present a potential tool for more quickly and thoroughly navigating the literature on expansive topics such as COVID-19.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The output of biomedical publication search tools is a set of manuscripts but access to content is restricted.

WHAT THIS STUDY ADDS

⇒ This study provides evaluation of a new method to extract, repurpose and disseminate information derived from published manuscripts.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Despite massive advancements in scientific methods, the science and practice of extraction and integration of information in manuscripts have lagged. We describe a method that advances a new method to improve efficiency of extraction and integration of information.

INTRODUCTION

PubMed contains over 32 000 000 publications and this collection grows by approximately 1 000 000 articles each year.[1 2] The first step in accessing this wealth of knowledge is to acquire a list of publications through search engines such as PubMed and Google Scholar. The user must then painstakingly comb through full-text articles to find the information they seek.

Community curation platforms such as Wikipedia allow rule-based descriptions of virtually any topic. However, Wikipedia is not designed to provide a comprehensive summary of the information contained within a set of publications. Community curation has been used on databases such as UniProt (model organism genome database) but these function more as annotation events than a system to extract information contained within full-text articles.[3] Artificial intelligence (AI) tools for information integration have sought to overcome these obstacles while simultaneously decreasing the time input required from the researcher. Many of the currently existing AI models are limited to generating detailed groupings of articles based on their contents or extracting and comparing information only within narrow categories.[4] These systems have yet to reach the point of providing users with thoroughly researched, discrete answers to their questions.

An online information management system (Refbin.com) was used to manually extract and integrate newly reported data from 1000 COVID-19 articles.[5] As reported in our recent publication,[5] extracted information was described using a minimum of four types of note fields (topic, population, description of the type of measurement and the actual reported measurement). Extracted results

in the same topic were merged so that parent topic note fields were shared. The full text of each article was read and individual observations, such as the incidence rate of COVID-19 infection in college students, were manually entered into the database. That piece of information was grouped with similar observations from other publications. Rather than organising and filing information based on the publication, this new COVID-19 publication database was organised into logical groups of data such as mental health issues related to lockdown or observations on maternal to fetal transmission of COVID-19. A user of this database can readily see specific sets of information and navigate rapidly to increasing levels of detail. In this database, the user does not need to necessarily know what they are looking for and perform a search. The user navigates through topics covering all of the data to quickly find the information they need.

The aim of this pilot trial was to determine whether this new type of database would provide an operational improvement over conventional methods to more rapidly and more accurately find evidence to support statements about COVID-19. A sample of students from the University of Vermont was randomly assigned to use the COVID-19 publication database or their usual preferred search methods to research eight prompts about COVID-19. If assigned to use their preferred research methods, students were allowed to use any means they knew of to find and read primary research articles. We report here the results showing an improved outcome using the new system.

METHODS

Study design

Eighteen participants were recruited from biomedically related courses and student organisations at the University of Vermont to participate in this randomised pilot trial. Prior to participating in the trial, the students completed a participant information survey including questions about their major, level of education, professional goals, research experience, usual speed of task completion and comfort with technology. The final sample included 17 students.

The participants were stratified based on self-rated questions about their previous research experience and usual task completion time requirements. Once stratified, the participants were randomly assigned to use either the new COVID-19 publication database (group A) or their preferred research methods (group B).

The trial was conducted over Microsoft Teams. Prior to taking the test, all participants received a 10 min step-by-step lesson on how to use the COVID-19 publication database prior to being assigned to group A or group B. The participants were informed that only evidence from primary literature would be counted as correct and that only 4 min were allowed per question. The tasks included finding specific pieces of information and were presented in survey form.

The relevance and accuracy of the evidence found for each prompt were individually graded by two members of our research team and then reviewed. Using parameters of specificity established during prompt development, each answer was graded in a binary of correct or incorrect. A correct answer submitted after 4 min (240 s) had passed was marked as incorrect. The time participants needed to answer each prompt was recorded by Qualtrics as the time elapsed between opening the question and clicking the submit button.

This pilot study qualified for an exemption from ethics review by the Institutional Review Board at the University of Vermont and the University of Vermont Medical Center. According to the definition of activities constituting research at these institutions, this pilot trial met the criteria for operational improvement activities and was, therefore, exempt from review.

Query prompt development

The prompts used in this investigation were written using a question stem stating a fact or observation about COVID-19, followed by a request for supporting evidence that the participant had to find using their assigned research method. At the time of this study, the COVID-19 database contained information on 1000 COVID-19 articles. Prompts were written by the study team that confirmed each prompt was answerable using both the COVID-19 database and Google (table 1). A preliminary investigation and refinement of the prompts were conducted with four students. Participants accessed the prompts through a Qualtrics survey, which randomly assigned them 8 of 18 total prompts.

Analysis of prompt and language difficulty was conducted using Microsoft Word's readability tool. The Flesch-Kincaid Grade Level test yielded a score of 13.4, meaning that the language used in these prompts was best suited for those with some college education.

Statistical analysis

Statistical analyses were conducted using SAS Version 9.4 and SYSTAT Version 11 software. Contingency tables were created using the number of correctly answered prompts and the variables: research method (COVID-19 Database or other methods), day of study participation,[1–17] academic background (major 1=biochemistry/biology; 2=dietetics, nutrition and food science (DNFS)), level of education (1=undergraduate; 2=graduate), previous research experience (1=uncomfortable; 2=neutral; 3=comfortable) and gender (0=male; 1=female).

A Cox proportional hazards' model was used to analyse these data in a time-to-correct answer context. The Cox model and survival analyses in general are used widely throughout medical literature but have had limited use in the computer science literature.[6–10] The event of interest in this study is a correctly answered question, enabling us to simultaneously evaluate completion time and accuracy. We examined the effect of several covariates on the time until a participant answered a prompt correctly or until their response was

Table 1 The complete set of prompts administered at random to participants

	Prompt
1	Find two primary research articles demonstrating that Black Americans and predominantly Black communities have suffered from higher incidences of SARS-CoV-2 than White Americans and predominantly White communities. One of these articles should present the data as a rate of infection, while the other should present evidence of this statement as an OR.
2	Find an article in which <3% of healthcare workers without symptoms of COVID-19 tested positive for SARS-CoV-2 through antibody testing. Find another article in which more than 18% of asymptomatic healthcare workers tested positive for SARS-CoV-2 through antibody testing.
3	Find two articles in which the amount of C reactive protein measured in critically severe COVID-19 patients was at least 1.5 times greater than that of severe COVID-19 patients.
4	Find one article showing that essential workers, people receiving treatment for PTSD, and young adults are among groups with the greatest risk of developing suicidal thoughts as a result of the pandemic. Report their risk in terms of ORs.
5	Reports of the percentage of newborns born to COVID-19 positive mothers who demonstrate fever symptoms are variable across the literature. Find one article in which none of the newborns had a fever and find another article in which more than 20% of the newborns had a fever. These articles cannot be case reports of only a single patient.
6	Several articles have published descriptions of new-onset psychotic disorder-like symptoms in patients with COVID-19 (meaning they had no prior history of these symptoms). Find an article that reports on the success of treatment for psychotic-like symptoms in patients with COVID-19. Your answer should include the percentage of patients for whom these symptoms were resolved.
7	Find an article that describes a new-onset of demyelinating lesions in the central nervous systems of patients with severe or critical COVID-19. The article you select must have a study population of at least 20 patients with COVID-19 positive.
8	Peak infectiousness typically occurs around 2 days before symptom onset. Find an article with a sample size of at least 1000 confirmed COVID-19 cases that describe this information and also describes the number of transmission events attributable to pre-symptomatic individuals.
9	Find one article that reports a range of recorded COVID-19 incubation periods with a maximum value of more than 18 days.
10	Find one article that reports the presence of SARS-CoV-2 in the sweat of symptomatic patients with COVID-19. The study should have a sample size of at least 25 individuals. Report the number of patients with COVID-19 with positive sweat samples.
11	Individuals with recent or lifetime substance abuse disorder (SUD) diagnoses are at higher risk of requiring hospitalisation and ventilation treatments after contracting COVID-19. They are also at a greater risk of dying due to the disease. Find an article that supports these three points using data presented as ORs. Note: these data should reflect increased risk as a result of SUD independently of any other comorbid conditions.
12	Find an article that describes a 70% or greater decrease in regional ICU admissions as a result of enacting physical distancing measures.
13	During the COVID-19 pandemic, a greater proportion of obese individuals have reported having a difficult time eating healthfully or are eating less healthy foods compared with the general population. Find two articles that support this statement. One article should report the proportion of obese individuals eating less healthfully and the other should report on the same measure in the general population.
14	Chest CT scans of hospitalised patients with COVID-19 show predominantly bilateral pulmonary lesions as opposed to only unilateral involvement. This trend is even more pronounced in patients admitted to the ICU. Find an article whose findings support this statement. Report the incidence of bilateral and unilateral pulmonary lesions in these two patient groups (hospitalised patients and patients admitted to the ICU).
15	There is evidence that SARS-CoV-2 can cross the blood-brain barrier to invade the brain. Find an article that reports on histochemical analysis of post-mortem brain tissue from patients with COVID-19 in which more than 95% of the samples studied showed evidence of astrogliosis (variably or in all brain regions).
16	The COVID-19 pandemic has had pronounced effects on mental health and substance abuse rates. Find an article in which 28% or more of the general population reported increased alcohol intake. These findings must also show a positive correlation with a history of mental illness.
17	Find two articles providing evidence that transmission of SARS-CoV-2 from mother to baby can occur through the placenta. One of these articles should provide evidence of transmission through immunostaining techniques, while the other should provide evidence using transmission electron microscopy.
18	Find an article showing that study participants who received the Moderna vaccine had a 0% incidence of severe COVID-19 infection post-vaccination.

ICU, intensive care unit; PTSD, post traumatic stress disorder.

censored for being incorrect or over time.[11] A sandwich estimator was used in conjunction with the Cox proportional hazards' model to account for the clustering of responses by each participant.[12] Kaplan-Meier plots were created to visualise the effect of the variables previously mentioned on the rate at which prompts were answered correctly and how many were answered correctly. The Kaplan-Meier estimator was selected due to its ability to handle right-censored data which took the form of wrong or incomplete answers in this pilot study.[13] A Cox proportional hazard model for the search method as well as bivariate models for demographic factors were examined.

RESULTS

Demographics and study groups

A total of 136 responses were collected, divided randomly, and nearly equally among 18 prompts. 72 responses were gathered from group A and 64 of these responses came from participants in group B (this discrepancy is due to incomplete data from the 18th participant) (table 2). One hundred and twenty of the total 136 responses were submitted by female participants reflecting the majority-female classes from which these students were recruited (online supplemental table 1). The study participants were stratified according to major and previous research experience and then randomly assigned to either group. This resulted in an equal number of the responses from group A (24) and group B (24) by students who identified as biology or biochemistry majors (online supplemental table 2). A similarly even distribution among the research methods was attained with students studying DNFS (group B=40, group A=48) (online supplemental table 2). The majority of the responses (104/136) were completed by participants who considered themselves

Table 2 Number of responses by research method group

Prompt number	Group A	Group B	Number of responses
1	4	2	6
2	3	5	8
3	2	5	7
4	5	4	9
5	4	2	6
6	5	3	8
7	7	2	9
8	3	3	6
9	4	4	8
10	4	3	7
11	2	6	8
12	4	4	8
13	4	4	8
14	4	3	7
15	6	2	8
16	4	4	8
17	3	4	7
18	4	4	8
Total	72	64	136

neither uncomfortable nor comfortable in working with biomedical literature (online supplemental table 3). This trial spanned 17 days, from April 26 to May 12, 2021, with a relatively even distribution in the date of study completion among the research method groups. Pearson Chi-square tests showed no association between any of the previously mentioned demographic factors and the research method groups to which the participants were assigned.

No significant associations were found using Pearson Chi-square when considering the day of participation, major and previous research experience in relation to the total number of correct responses across both research method groups. Within group A, biochemistry and biology majors answered 62.5% of their prompts correctly and DNFS students answered 45.8% correctly (p=0.18) (table 3). When looking at these same responses through the lens of previous research experience, it was found that group A students who rated themselves in the highest available category for comfort reading primary literature had a correct response rate of 75%, while the eight students who rated themselves in the middle category had a 48.4% accuracy rate (p=0.15) (table 4). Among

Table 3 Percentage of correct responses from group A participants by major

	Incorrect	Correct	N
Biology or biochemistry	37.5	62.5	24
Dietetics, nutrition and food sciences	54.2	45.8	48
Total	48.6	51.4	
N	35	37	72

Table 4 Percentage of correct responses from group A participants by comfort using and reading research literature (research experience)

	Incorrect	Correct	N
Neutral	51.6	48.4	64
Comfortable	25	75	8
Total	48.6	51.4	
N	35	37	72

the correct and incorrect responses obtained by group B, none of the evaluated demographic factors approached significance.

Kaplan-Meier plots: combined and by research method

Kaplan-Meier plots were used to simultaneously examine the accuracy of each response and the time required to obtain it. Using the survival analysis terminology frequently found in biomedical literature, exact failures or deaths correspond to prompts answered correctly and survival time represents the number of seconds required for the correct response. Right censored responses are those that were either obtained in more than 240 s or were incorrect. The survivor function in the Kaplan-Meier plots below, therefore, demonstrates the proportion of responses that were correct (died) within the time spent on each response.

The Kaplan-Meier plot in figure 1 shows the prompts answered by group A were more frequently answered correctly and more rapidly than group B. A log-rank test yielded a p value <0.001. The mean survival times for prompts at risk of failing (being answered correctly) were 226.8 s and 365.5 s for group A and group B, respectively. The overlap in the survival curves in figure 1B indicates a degree of homogeneity among the group A participants (p=0.259). Figure 1c contains the survival plot of group B. The overlap of curves here also indicates a degree of homogeneity among group B (p=0.466). Due to there only being 12 correct responses out of 64 at risk in group B, both the similarity of the survival curves and the mean survival time could be slightly distorted. This was not an issue in group A where there was a total of 37 correct responses out of 72 prompts answered.

Cox proportional hazards' models with Sandwich Estimator

A Cox proportional hazard model using a Sandwich Estimator was conducted as described by Lin and Wei to account for the clustering of responses by participant.[12] The analysis of the research method resulted in a parameter estimate for group A of 1.388 (SE of 0.254, p<0.001). The HR of obtaining a correct response in the context of the method parameter was 4.01 (95% CI 2.433 to 6.579).

Bivariate Cox models of analyses for the research method and gender, day of study participation or major demonstrated a consistently more significant influence of the research method on response accuracy than the alternate variable. The bivariate model parameter estimate for

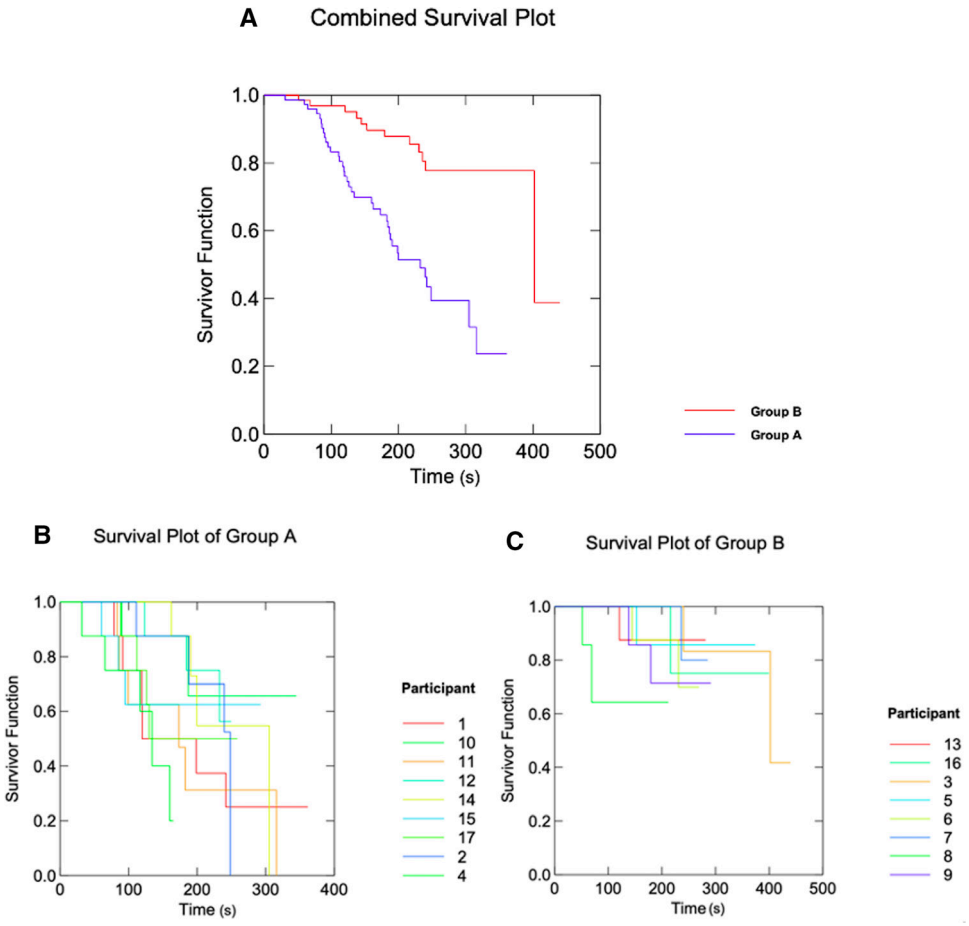


Figure 1 Kaplan-Meier plots showing the survivor function for group A compared with group B (A), as well as for the participants within group A (B) and group B (C) independently.

women was 0.144 (SE of 0.168, p=0.39), for day of study participation −0.008 (SE of 0.014, p=0.579) and for major 0.392 (SE of 0.235, p=0.096). The bivariate hazard ratios were as follows: woman, 1.154 (95% CI 0.830 to 1.606); study day, 0.993 (95% CI 0.67 to 1.019); and major, 0.676 (95% CI 0.426 to 1.072). The level of research experience, in contrast to the previous parameters, resulted in a more significant parameter estimate of 0.542 (SE of 0.118, p<0.0001) and a HR of 1.719 (95% CI 1.363 to 2.167). A correlation matrix was used to show that despite the larger parameter estimate of previous research experience, research method and experience were independent (r=0.022).

DISCUSSION

This randomised pilot trial was performed to determine whether a new information platform containing information from 1000 COVID-19 publications enabled faster and more accurate answers to prompts than conventional methods of accessing biomedical information. Kaplan-Meier plots and Cox Proportional Hazard Models confirmed that the new method of information integration (group A) enabled participants to answer prompts about COVID-19 more quickly and accurately.

Tests for respondent heterogeneity within groups A and B were negative. The paucity of correct answers collected by group B may in part be attributable to the difference between the volume of sources retrieved by a Google search compared with a search in the group A COVID-19 database. After completing the prompts, many of the members in group B commented that 4 min was simply not enough time to find answers. This rush for time coupled with the many blank responses seen from this group could reflect the negative correlation between time per question and user satisfaction described by Xu and Mease.[14] The discrepancy in timed-out prompts and blank responses between the two groups could simply be the result of the pace at which research was completed using the two methods, but it could also provide insight into frustration among participants using traditional research methods. Allowing each participant to be exposed to the full range of questions was part of the design for comparability among students. A limit of 4 min allowed each participant to be exposed to the full range of questions. This was a study design tradeoff between comparable exposure within a specific time frame versus unbounded time to complete the search and having a non-comparable opportunity to answer all specific questions. The choices of the questions were also designed to be not that sophisticated

as to make questions too difficult to be able to abstract from each of the differing approaches. There are differences when conducting real-world research and this pilot evaluation study of the interface and database.

A significant difference was seen between the number of correct answers gathered by group A and group B. This observation contrasts with other examples from the literature.[15 16] For example, when using a question answering system compared with a document retrieval system, Smucker *et al* found that participants answered questions correctly at similar rates.[15] The question answering system described in the Smucker study resembles the new COVID-19 database in that they both aim to present condensed, succinct information independently of its source articles. Therefore, we suspect that the marked improvement in answer accuracy observed in our study may be related to the format in which information is presented. The parent–child organisation structure in the new database system may enable users to more easily adjust the information they are viewing without repeatedly editing their query.

A set of bivariate Cox proportional hazard models showed that among the demographic factors considered in this pilot study previous research experience was the only one to confer an advantage in study performance. A correlation matrix, however, indicated that the prestratification of participants had evenly distributed their degree of prior research experience between the two groups. When examining responses from just group A, major (biology and biochemistry v. dietetics, nutrition and food sciences) and previous research experience did approach significance. This indicates that familiarity with research literature and biomedical language is potentially advantageous in using the database, but ultimately not to a significant extent. In general, studies evaluating information retrieval systems with user-oriented methods have had similar trouble identifying significant extraneous influences on user performance. Whether collecting data on age, sex, computer experience, online search engine familiarity or career objectives these examples from the literature show no significant differences between the groups assigned to the different search methods nor accuracy rates.[16–18]

In this pilot randomised trial, statistically significant differences were observed between groups A and B. This emphasised the marked differences between the methods and warrant expansion of the trial to validate the results. In addition to increasing the sample size, it will be helpful to source queries from a separate set of participants, experts on the topics addressed by the database or search history data from commercial engines.[14 16 19] It will also be useful to expand the diversity of participants' education status. Determining the range of students and users that could benefit from this new platform for information will help guide integration into educational systems. A limitation of the Refbin system is that, unlike Google Scholar and PubMed, methods to continuously update the database are not yet implemented.

CONCLUSION

We demonstrated that a new method of extracting information from published biomedical literature allows users to more quickly and more accurately answer questions related to COVID-19. Topics such as COVID-19 present so much data that a next-generation system is required to more rapidly allow users to answer questions related to the published data. This pilot study with 1000 COVID-19 manuscripts points to a possible solution to speed up research.

Contributors CTL is guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests David Krag has a significant financial interest in Plomics Inc, the developer of RefBin.com. The investigator disclosed his personal financial interest to the IRB of the University of Vermont. The IRB at the University of Vermont and the University of Vermont Medical Center determined that this pilot study qualified for an exemption from ethics review. Any potential conflicts of interest were managed. All other authors declare that they have no conflicts of interest.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants but all methods were carried out in accordance with relevant guidelines and regulations as defined by the University of Vermont Institutional Review Boards (IRBs) serving the University of Vermont and the University of Vermont Medical Center. The present pilot study was classified as operational improvement activities by the IRBs. According to the definition of activities constituting research at these institutions, the methods of this pilot did not qualify as research and were exempt from ethics review. The IRBs thereby approved all methods carried out in this pilot study to proceed without further review. Informed consent was obtained from all individual participants included in this pilot study, exempted this study. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iD

David Krag http://orcid.org/0000-0001-5355-5999

REFERENCES

1 Simon C, Davidsen K, Hansen C, *et al*. BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics* 2019;19:57.
2 Müller H-M, Van Auken KM, Li Y, *et al*. Textpresso central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. *BMC Bioinformatics* 2018;19:94.
3 LudÃ¤scher B, Lin K, Bowers S. Managing scientific data: From data integration to scientific workflows. In: *Gsa today (special issue on Geoinformatics*. 109, 2006.
4 Extance A. How AI technology can tame the scientific literature. *Nature* 2018;561:273–4.

5 Lunna S, Flinn I, Prytherch J, *et al*. 'Refbin' an online platform to extract and classify large-scale information: a pilot study of COVID-19 related papers. *BMJ Health Care Inform* 2022;29:e100452.
6 Ortega F, Convertino G, Zancanaro M. Assessing the Performance of Question-and-Answer Communities Using Survival Analysis. ArXiv14075903 Cs [Internet], 2014. Available: http://arxiv.org/abs/1407.5903 [Accessed 08 Jun 2021].
7 Ortega F, Izquierdo-Cortazar D. Survival analysis in open development projects. In: *2009 ICSE workshop on emerging trends in Free/Libre/Open source software research and development*, 2009: 7–12.
8 Zhang D, Prior K, Levene M. How long do Wikipedia editors keep active? In: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration - WikiSym '12 [Internet]. Linz, Austria: ACM Press, 2012. Available: http://dl.acm.org/citation.cfm?doid=2462932.2462938 [Accessed 08 Jun 2021].
9 Lam SK, Uduwage A, Dong Z. WP:clubhouse? An exploration of Wikipedia's gender imbalance, 2011. Available: https://experts.umn.edu/en/publications/wpclubhouse-an-exploration-of-wikipedias-gender-imbalance [Accessed 08 Jun 2021].
10 Samoladas I, Angelis L, Stamelos I. Survival analysis on the duration of open source projects. *Inf Softw Technol* 2010;52:902–22.
11 Finch H, Lapsley D, Baker-Boudissa M. A survival analysis of student mobility and retention in Indiana charter schools. *Educ Policy Anal Arch* 2009;17:18.
12 Lin DY, Wei LJ. The robust inference for the COX proportional hazards model. *J Am Stat Assoc* 1989;84:1074–8.
13 McNeish D. Applying Kaplan-Meier to item response data. *The Journal of Experimental Education* 2018;86:308–24.
14 Xu Y, Mease D. Evaluating web search using task completion time. In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '09 Internet*. Boston, MA, USA: ACM Press, 2009: 676. http://portal.acm.org/citation.cfm?doid=1571941.1572073
15 Smucker MD, Allan J, Dachev B. Human question answering performance using an interactive information retrieval system 2012;9.
16 Hersh W, Pentecost J, Hickam D. A task-oriented approach to information retrieval evaluation. *J Am Soc Inf Sci* 1996;47:50–6.
17 Al-Maskari A, Sanderson M, Clough P. The Good and the Bad System: Does the Test Collection Predict Users' Effectiveness? 2008:59–66.
18 Egan DE, Remde JR, Gomez LM, *et al*. Formative design evaluation of superbook. *ACM Trans Inf Syst* 1989;7:30–57.
19 Lewandowski D. Evaluating the retrieval effectiveness of web search Engines using a representative query sample. *J Assoc Inf Sci Technol* 2014;66.

BMJ Health & Care Informatics

Clinical utility of automatic phenotype annotation in unstructured clinical notes: intensive care unit use

Jingqing Zhang , Luis Daniel Bolanos Trujillo , Ashwani Tanwar , Julia Ive , Vibhor Gupta, Yike Guo

To cite: Zhang J, Bolanos Trujillo LD, Tanwar A, *et al*. Clinical utility of automatic phenotype annotation in unstructured clinical notes: intensive care unit use. *BMJ Health Care Inform* 2022;**29**:e100519. doi:10.1136/bmjhci-2021-100519

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2021-100519).

JZ, LDBT and AT contributed equally.

Received 29 November 2021
Accepted 30 September 2022

Pangaea Data Limited, London, UK

Correspondence to
Professor Yike Guo;
yguo@pangaeadata.ai

ABSTRACT

Objective Clinical notes contain information that has not been documented elsewhere, including responses to treatment and clinical findings, which are crucial for predicting key outcomes in patients in acute care. In this study, we propose the automatic annotation of phenotypes from clinical notes as a method to capture essential information to predict outcomes in the intensive care unit (ICU). This information is complementary to typically used vital signs and laboratory test results.

Methods In this study, we developed a novel phenotype annotation model to extract the phenotypical features of patients, which were then used as input features of predictive models to predict ICU patient outcomes. We demonstrated and validated this approach by conducting experiments on three ICU prediction tasks, including in-hospital mortality, physiological decompensation and length of stay (LOS) for over 24 000 patients using the Medical Information Mart for Intensive Care (MIMIC-III) dataset.

Results The predictive models incorporating phenotypical information achieved 0.845 (area under the curve–receiver operating characteristic (AUC-ROC)) for in-hospital mortality, 0.839 (AUC-ROC) for physiological decompensation and 0.430 (kappa) for LOS, all of which consistently outperformed the baseline models using only vital signs and laboratory test results. Moreover, we conducted a thorough interpretability study showing that phenotypes provide valuable insights at both the patient and cohort levels.

Conclusion The proposed approach demonstrates that phenotypical information complements traditionally used vital signs and laboratory test results and significantly improves the accuracy of outcome prediction in the ICU.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previous studies have demonstrated exemplary performance for predicting outcomes in the intensive care unit (ICU) using bedside measurements and laboratory test results.
⇒ Contextual embeddings from recent transformer-based natural language processing models have enabled more accurate detection of medical concepts.

WHAT THIS STUDY ADDS

⇒ This study introduces a new methodology that incorporates contextualised phenotypical features from clinical texts and their persistence in modelling ICU time-series prediction tasks. The persistence of phenotypes in this study refers to the frequency with which each phenotype appeared in each patient record.
⇒ We conducted an interpretability study to illustrate why and how phenotypical features are highly relevant to ICU outcome prediction.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Researchers should explore how other clinical cases can benefit from the use of phenotypical information.
⇒ Given the richness of the information found in the clinical text, all parties involved should work together to enable secure and ethical use.

INTRODUCTION

Unprecedented amounts of healthcare data have been accumulated. The National Health Service (NHS) datasets record billions of patient interactions annually.[1] In an intensive care unit (ICU), a wealth of data is generated for each patient owing to the necessity of closely monitoring of patients,[2] with new vital information recorded every minute.

Conventionally, electronic health records (EHRs) contain two data types. These are structured (blood tests and temperature) and unstructured data (radiology reports and discharge summaries), with the latter comprising the most extensive part of patient data (up to 80%[3]). Both types of data are valuable for ICU monitoring. The majority of recent research[4–6] relies heavily on analysable structured data—typically laboratory test results and vital signs.

Among the clinical features found in unstructured data, phenotypes [In the medical text, the word 'phenotype' refers to deviations from normal morphology, physiology or behaviour[7]]. have received the least attention in ICU monitoring.[8] This is mainly because of the challenge of extracting phenotypical information expressed in different

contexts and contextual synonyms. For example, the phenotype *hypotension* can be expressed as 'a drop in blood pressure' or 'BP of 79/48'. However, phenotypical information is crucial for understanding disease trajectory, identifying important disease-specific information, stratifying patients and identifying novel disease subtypes.[9]

This study thoroughly investigated the value of phenotypical information extracted from textual clinical data in ICU monitoring. Furthermore, this study expands on our previous study[4] by incorporating additional clinical data from free-text clinical notes instead of limiting the patients' profiles to data recorded in a structured format. Based on our previous study,[10] we automatically extracted phenotypes from clinical texts using a self-supervised methodology. This methodology includes contextualised word embeddings,[11] which are particularly helpful in detecting contextual synonyms. We extracted over 15 000 phenotypical concepts of the human phenotype ontology (HPO).[12] Phenotypical features extracted in this manner were enriched with information from the structured data. To interpret our results, we used Shapley additive explanation (SHAP) values.[13].

We benchmarked our approach on the following three mainstream ICU tasks following practice[4] for comparison: length of stay (LOS), in-hospital mortality and physiological decompensation.

Our main contributions are (1) the approach to incorporating phenotypical features from clinical textual data into the modelling of ICU time-series prediction tasks, (2) investigation of the importance of the phenotypical features in unstructured clinical data in combination with structured information for predicting patient course at the microlevel (individual patient) and macrolevel (cohort), (3) a thorough interpretability study demonstrating the importance of phenotypical and structured features for ICU cases, and (4) demonstration of the utility of automatic phenotyping for ICU use cases.

METHODOLOGY

Data preprocessing

The MIMIC-III, a publicly available ICU database, was used in this study.[14] MIMIC-III contains 49 785 anonymised hospital admissions and 38 597 distinct adult patients (aged 16 years or above) admitted to a critical care unit (eg, coronary care unit, cardiac surgery recovery unit, medical ICU, surgical ICU and trauma surgical ICU) of the Beth Israel Deaconess Medical Center between 2001 and 2012. The median age of the adult patients was 65.8 years (Q1–Q3: 52.8–77.8); 55.9% of patients were male; and diseases of the circulatory system were the most prevalent (based on ICD-9: 390–459), accounting for 36.6% of all adult admissions.

We followed the standard practice[4] of defining the three ICU tasks, data collection and data preprocessing on MIMIC-III. First, we formulated the *in-hospital mortality* problem as a binary classification 48 hours after admission, in which the label indicates whether the patient died before discharge. Second, *physiological decompensation* was formulated as a binary classification in which the target label corresponds to whether the patient will die in the next 24 hours. Finally, we cast the *LOS* prediction task as a multiclass classification problem, where the labels correspond to the remaining LOS. Possible values were divided into 10 bins: 1 for a LOS of less than a day, 7 bins for each day of the first week, another bin for the LOS of more than a week but less than 2 weeks, and the final bin for stays of more than 2 weeks.

For data collection, we used structured and unstructured data following the filtering criteria[4] for patients in all three tasks. In addition, we discarded all ICU episodes in which a clinical note had not been recorded. This reduced our training and test data compared with the benchmark[4]; therefore, we recalculated the baseline scores using their code on our new test set for a fair comparison. Overall, there were over 24 000 patients, and the exact numbers of patients, ICU episodes and time steps per task are reported in section A.1 of the online supplemental appendix. The mortality rate across all patients was 13.12%, and the decompensation rate across all time steps was 2.01%. Most patients stayed in the ICU for less than 7 days, and the distribution of ICU admissions per LOS class is presented in detail in section A.2 of the online supplemental appendix.

For the preprocessing of structured data, we followed steps [Accessed in November 2021 (https://github.com/YerevaNN/mimic3-benchmarks)] to collect 17 clinical features (capillary refill rate, diastolic blood pressure, fraction inspired oxygen, Glasgow Coma Scale eye opening, Glasgow Coma Scale motor response, Glasgow Coma Scale verbal response, Glasgow Coma Scale total glucose, heart rate, height, mean blood pressure, oxygen saturation, respiratory rate, systolic blood pressure, temperature, weight and pH). Highly specific features may not be routinely obtained for all patients. The inclusion of such features renders the algorithm less generalisable. Therefore, additional structured clinical features were excluded to ensure comparability with the previous studies. For data preprocessing of unstructured data, we collected all free-text clinical notes, including nursing notes, physician notes and discharge summaries, at all time steps during the ICU stay. High data sparsity was observed because clinical notes were recorded roughly every 12 hours. Processed structured and unstructured data were used as inputs for the proposed approach.

Algorithm development and analysis

The proposed approach consisted of two steps. The first step collected clinical features (specifically, phenotypical features standardised by HPO[12]) from unstructured data using natural language processing algorithms. The second step combined the phenotypical features from unstructured data and the 17 clinical features from structured data as input features for machine learning classifiers to separately predict in-hospital mortality, physiological

decompensation and LOS. Unlike the study by Harutyunyan *et al*,[4] which used only 17 features from structured data as input, the proposed approach combines clinical features from unstructured and structured data.

First, we applied a state-of-the-art phenotyping model to extract phenotypes from free-text clinical notes.[10] This model leveraged self-supervised pretraining techniques, contextualised word embeddings by the transformer model[15] and data augmentation techniques (paraphrasing and synthetic text generation) to capture names (such as hypertension), synonyms (such as high blood pressure), abbreviations (such as HTN) and, more importantly, contextual synonyms. For example, descriptive phrases such as 'rise in blood pressure', 'blood pressure is increasing' and 'BP of 140/90' are considered contextual synonyms of hypertension (HP:0000822), and finding such contextual synonyms requires an understanding of the context. Thus, the phenotyping model demonstrated superior performance to alternative algorithms. We refer readers to a recent work[10] for the methodological details. For comparison, we used alternative phenotyping methods, including ClinicalBERT[11] (fine-tuned for phenotyping) and neural concept recogniser (NCR).[16] NCR uses convolutional neural networks to assign similarities to HPO concepts encoded by pretrained non-contextualised word embeddings.

Second, the phenotypical features were combined with structured clinical features as input features for machine learning classifiers to predict the three ICU tasks. We used standard machine learning classifiers: random forest (RF)[17] and long short-term memory (LSTM) networks[18] for prediction (the hyperparameters are available in section A.3, online supplemental appendix). We distinguished between persistent and transient phenotypical features to reduce sparsity.

More precisely, phenotypes that were clinically likely to last an entire admission in the vast majority of typical cases (eg, tuberculosis and cancer) were labelled as 'persistent'. Phenotypes acquired or improved during a stay, such as fever or cough, were labelled as 'transient'. We developed transient and persistent phenotypes from the moment they appeared until a new clinical note appeared or until the end of the stay. Phenotype transiency and persistency patterns were leveraged in the model to make clinical predictions. This was found to be beneficial and will be discussed in the Phenotype persistency section. We also addressed data sparsity by aggregating the HPO terms into their parents (according to the HPO hierarchy).

Evaluation metrics

For comparison with Harutyunyan *et al*,[4] we used area under the curve–receiver operating characteristic (AUC-ROC)[19] and area under the curve-precision recall (AUC-PR) for in-hospital mortality and physiological decompensation tasks. We primarily relied on AUC-ROC for statistical analysis because it is threshold independent and used by the benchmark[4] as the primary metric. For the LOS task, we used Cohen's kappa[20] and the mean absolute deviation,[21] with emphasis on the kappa scores for statistical analysis.

Model evaluation and statistical analysis

A train-test split based on the benchmark was used,[4] with patients without clinical notes excluded, resulting in 21 346 and 3824 patients for the training and test sets, respectively. Further, we performed fourfold cross-validation on the training set, and all splits were deterministic. We used bootstrap resampling following the benchmark for statistical analysis. To compute CIs on the test set, we resampled it 1000 times for LOS and decompensation, and 10 000 times for in-hospital mortality tasks. We then computed the scores on the resampled data for the 95% CIs.

To provide interpretability for model predictions, we used SHAP values,[13] the implementation details of which are explained in section A.4 (online supplemental appendix). The SHAP values were used to explain the black box models and allowed us to quantify feature importance and their positive or negative impact on the outcome.

RESULTS

The performance of two classifiers was investigated: RF and LSTM. For each, we investigated the following features: structured features only (S) and structured features enriched with phenotypical features from one of the three phenotype annotators (ClinicalBERT, NCR and Ours). Across all tasks, ClinicalBERT found 664 phenotypes; NCR found 1441 phenotypes, and our model found 1446 phenotypes, on average. Thirty per cent of these phenotypes are persistent.

The results are presented in table 1, and those from the statistical tests are presented in section A.5 (online supplemental appendix). Additionally, we provide the cross-validation results in section A.6 (online supplemental appendix). Overall, they showed that phenotypical information positively complements structured information to improve performance on all tasks. The improvements with our phenotyping model were statistically significant across all tasks compared with using S or alternative phenotyping algorithms, except for in-hospital mortality with RF. This is explained by the fact that phenotypes carry highly valuable information.[22 23]

DISCUSSION

Decision support systems in the healthcare domain should be reliable, interpretable and robust; therefore, we accompanied the aforementioned results with a thorough study on interpretability both at the patient and cohort levels and an assessment of robustness by studying disease-specific subcohorts.

Table 1 Results for (A) in-hospital mortality, (B) physiological decompensation and (C) LOS

(A) In-hospital mortality			
Classification model	**Features design**	**AUC-ROC ↑**	**AUC-PR ↑**
SAPS-II	–	0.756	0.312
APACHE-III	–	0.733	0.308
RF	S	0.800 (0.775 to 0.824)	0.339 (0.286 to 0.395)
	S+NCR	0.828 (0.802 to 0.853)	0.467 (0.404 to 0.529)
	S+CB	0.812 (0.787 to 0.838)	0.403 (0.345 to 0.463)
	S+Ours	0.845 (0.826 to 0.873)	0.462 (0.404 to 0.524)
LSTM	S^4	0.825	0.41
	S	0.826 (0.801 to 0.848)	0.391 (0.334 to 0.452)
	S+NCR	0.841 (0.818 to 0.864)	0.453 (0.393 to 0.513)
	S+CB	0.826 (0.802 to 0.849)	0.414 (0.355 to 0.476)
	S+Ours	0.845 (0.823 to 0.868)	0.464 (0.405 to 0.523)
(B) Physiological decompensation			
Classification model	**Features design**	**AUC-ROC ↑**	**AUC-PR ↓**
RF	S	0.826 (0.821 to 0.831)	0.130 (0.123 to 0.138)
	S+NCR	0.825 (0.821 to 0.830)	0.124 (0.118 to 0.131)
	S+CB	0.826 (0.821 to 0.830)	0.125 (0.118 to 0.132)
	S+Ours	0.845 (0.840 to 0.850)	0.180 (0.171 to 0.190)
LSTM	S^4	0.809	0.125
	S	0.824 (0.819 to 0.829)	0.126 (0.119 to 0.133)
	S+NCR	0.834 (0.829 to 0.839)	0.134 (0.127 to 0.142)
	S+CB	0.833 (0.828 to 0.838)	0.114 (0.108 to 0.119)
	S+Ours	0.839 (0.834 to 0.844)	0.145 (0.138 to 0.153)
(C) LOS			
Classification model	**Features design**	**Kappa ↑**	**MAD ↓**
RF	S	0.390 (0.388 to 0.392)	136.8 (136.2 to 137.4)
	S+NCR	0.390 (0.388 to 0.392)	142.5 (141.9 to 143.1)
	S+CB	0.376 (0.374 to 0.379)	144.3 (143.7 to 144.9)
	S+Ours	0.420 (0.418 to 0.422)	110.3 (109.3 to 111.3)
LSTM	S^4	0.395	126.7
	S	0.380 (0.377 to 0.382)	157.0 (156.3 to 157.6)
	S+NCR	0.406 (0.404 to 0.408)	123.3 (122.8 to 123.9)
	S+CB	0.388 (0.386 to 0.390)	120.1 (119.6 to 120.6)
	S+Ours	0.430 (0.427 to 0.432)	116.7 (116.1 to 117.2)

Test set scores are shown with 95% CIs in brackets, if applicable. The best score for each classifier is highlighted in bold. The first row of LSTM refers to scores reported in previous literature, whereas the second row regards scores reproduced in this study with a comparable cohort. Here, S refers to structured; NCR refers to neural concept recogniser[16]; CB refers to ClinicalBERT; and Ours refers to our phenotyping model.

APACHE-III, Acute Physiology and Chronic Health Evaluation III; AUC-ROC, area under the curve–receiver operating characteristic; CB, ClinicalBERT; LOS, length of stay; LSTM, long short-term memory; MAD, mean absolute deviation; RF, random forest; SAPS-II, Simplified Acute Physiology Score II.

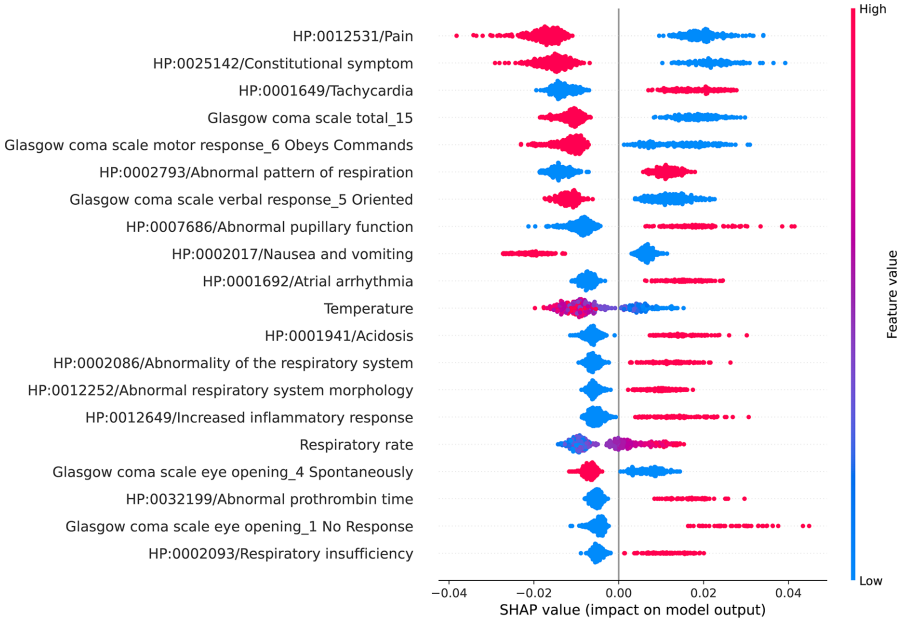
Phenotype persistency

We determined that it was beneficial to propagate phenotypes forward in time. Each phenotype is marked by a human clinical expert based on whether it typically persists throughout the ICU stay. Consequently, transient (eg, fever, cough and dyspnoea) and persistent (eg, diabetes and cancer) phenotypes propagate until the appearance of a new clinical note or the end of stay, respectively. We performed an ablation study and observed that phenotype propagation was more beneficial for RF than LSTM. The RF models with phenotype propagation achieved 4.6% higher AUC-ROC for in-hospital mortality, 2.5%

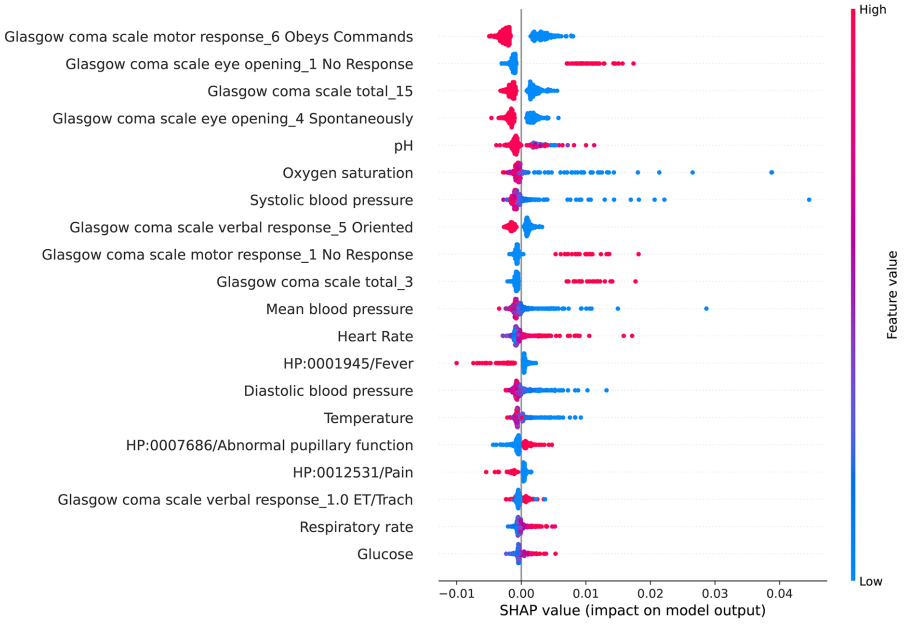
higher AUC-ROC for decompensation and 3.4% higher kappa for LOS than RF without phenotype propagation. However, LSTM with phenotype propagation achieved 1.4% higher AUC-ROC for in-hospital mortality, comparable results for decompensation, and 1.1% lower kappa for LOS. We hypothesise that LSTM, by design, can better capture temporal relationships, given a large amount of data to learn from. The results are presented in section A.7 (online supplemental appendix).

Phenotype importance

To elucidate the contribution of phenotypical features to prediction performance, the most important features were studied using the SHAP values.[13] Because computing



A In-hospital Mortality



B Physiological Decompensation

Figure 1 Top features for in-hospital mortality and physiological decompensation. Features are sorted in decreasing importance according to their mean absolute SHAP values. Each row presents a condensed summary of the feature's impact on the prediction. Each data sample is represented as a single dot in each row, and its colour on a particular row represents the value of that sample for that feature, with blue corresponding to lower values or absence and red corresponding to higher ones or presence.The SHAP value (horizontal position of a dot) measures the contribution of that feature on a sample, towards the prediction (right corresponding to mortality or decompensation and left corresponding to survival or out of decompensation risk). For instance (in A, top row), since the vertical axes clearly splits patients by colour, manifesting HP:0012531 pain consistently leads to lower chances of dying. SHAP, Shapley additive explanation.

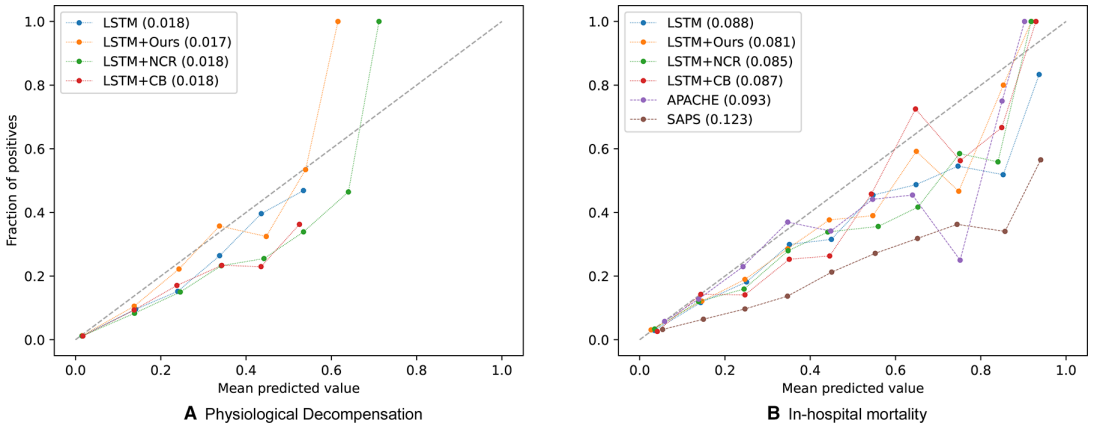


Figure 2 Calibration curves with LSTM for (A) physiological decompensation and (B) in-hospital mortality. Calibration curves are presented with its Brier score (the lower the better). Note that overall inclusion of phenotypical features from unstructured data helps with calibration. LSTM in legend refers to using structured features only. Ours, NCR, CB: phenotypical features from our phenotyping model, neural concept recogniser and ClinicalBERT, respectively. LSTM, long short-term memory; SAPS, Simplified Acute Physiology Score.

SHAP values is computationally complex, to accelerate the computation, this and all subsequent interpretability analyses based on SHAP values were conducted on the RF models rather than on the LSTM models. An illustration of our investigation is shown in figure 1, where we present the top predictive features for in-hospital mortality and physiological decompensation. This confirms that phenotypical features are beneficial for in-hospital mortality prediction, given that 13 of the 20 most important features are phenotypes. This is explained by the fact that forecasts need to rely on information that can provide insights accurately into the long-term future.

Contrary to bedside measurements, which may not correlate well with future outcomes owing to their dynamic nature, phenotypes are highly informative, given that they capture, for instance, comorbidities, which are essential for predicting mortality.[23] Furthermore, another study[24] including 230 000 ICU patients found that combining comorbidities with acute physiological measurements yielded the best results, outperforming traditional mortality scores (APACHE-II[25] and SAPS-II[26]).

Interestingly, the top-ranking feature for mortality prediction is whether the patient experiences pain. We also observed that the second top-ranking feature is *constitutional symptoms (HP:0025142)*. Noting that this is actually the resulting phenotype after aggregating all of its children, this phenotype should be interpreted not as a textual mention in the patient's EHR of the broad term but rather as a mention of any of its children (most

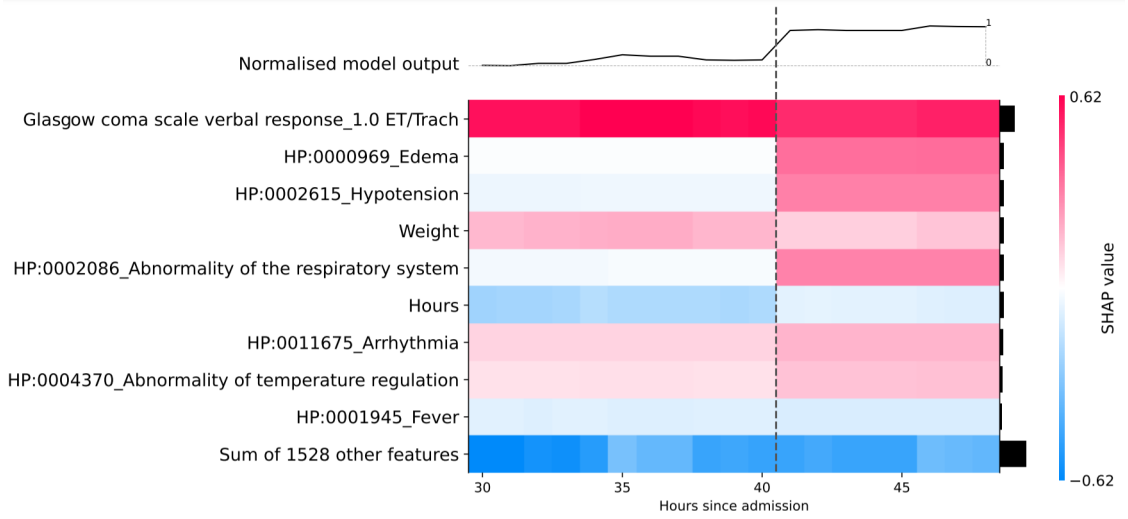


Figure 3 Illustrative case for an ICU length of stay of more than 14 days. Time course of the normalised predicted probability for a stay of more than 14 days and feature heatmap for a representative segment of the ICU stay. Each row of the heatmap represents one of the top features. At each time step, a feature can contribute positively (red) or negatively (blue) for predicting a stay of 14 days or more. Black horizontal bars at the right of each row represent the importance of the features. Note that a new clinical note that is available at the 42nd hour (vertical dashed line) leads to an increase in confidence of longer stay due to new features. Given the appearance of oedema, hypotension and abnormality of the respiratory system, the probability of a long stay increases from 69% to 88%. ICU, intensive care unit.

notably generalised pain). Consequently, the second feature again highlights the importance of pain.

Although not decisive, some initial evidence corroborates the fact that pain management improves outcomes in the ICU.[27] However, pain can be interpreted as a proxy for establishing a high level of consciousness, which has been correlated with better outcomes in the ICU.[28]

The other top-ranking phenotypes, such as atrial arrhythmia, nausea and vomiting, cover most of the body systems (ie, heart, lungs, gastrointestinal tract, central nervous system, coagulation, infection and kidneys), which are typically assessed using clinically validated scores, for example, APACHE-II and SAPS-II.

Our study also showed that, although phenotypical features are not as important for decompensation as for in-hospital mortality (only 3 out of the top 20 features for this task were phenotypes), they are nonetheless useful because they provide a more accurate estimation of the predicted risk. Given that this task is concerned with predicting mortality within the next 24 hours, bedside measurements become more informative because of their temporal correlation (illustrated in section A.8, online supplemental appendix). Nevertheless, bedside measurements can be ambiguous or provide an incomplete picture of the patient's status without the data found in clinical notes. For example, for one patient, *neoplasm of the respiratory system (HP:0100606)* was found to be the top feature, and although this phenotype was persistent, it increased the risk of decompensation appropriately, providing an overall better estimation. An illustration of this patient is presented in section A.9 (online supplemental appendix).

Similarly, the top features for long lengths of stay (more than 1 week) are presented in section A.10 (online supplemental appendix), where 10 of the 20 top features are phenotypes.

Calibration

The calibration of machine learning models compares the distribution of the probability predicted by the models with the distribution of probabilities observed in real patients. To measure model calibration, we used the Brier score[29] (the lower, the better). Our investigation of the respective calibration curves (see figure 2 and section A.11, online supplemental appendix) shows that phenotypes from unstructured notes improve model calibration across set-ups, especially for physiological decompensation and in-hospital mortality, which means that the distribution predicted by the models is closer to the real distribution of patients.

Prognosis analysis

Beyond producing clinically relevant explanations at the cohort level, with the help of SHAP values, we characterised a patient's disease trajectory and retrospectively discovered when and why the patient was the most vulnerable. For example, the fragment of a patient's LOS forecast in figure 3 illustrates an estimated probability, 41 hours after admission, of an LOS longer than 14 d of 69%, mainly because the patient scored one on the Glasgow Coma Scale verbal response. One hour later, when a clinical note became available, worrisome phenotypes appeared (including oedema, hypotension and abnormality of the respiratory system). Consequently, the estimated probability increased to 88%.

Cohort study

To understand its robustness, the performance of the proposed approach was assessed on cohorts of patients with different diseases, especially under-represented diseases. The test set was split into four disease-specific cohorts for patients with cardiovascular disease, diabetes, cancer and depression. The accuracies of the best LSTM models (using structured and phenotypical features) were reported individually for each cohort in each ICU task.

For in-hospital mortality and physiological decompensation, we observed comparable accuracies across the four cohorts. We reported an AUC-ROC range of 0.780–0.826 for in-hospital mortality and 0.792–0.820 for physiological decompensation in the four cohorts. In contrast, for LOS, we observed lower kappa values of 0.321 and 0.330 for small cohorts with cancer and depression, respectively, as opposed to 0.413 and 0.424 for larger cohorts with cardiovascular diseases and diabetes. We hypothesised that the nature of diseases has strong implications for in-hospital mortality and physiological decompensation, whereas LOS can be influenced by more factors that require larger data samples to model their interactions. The results are presented in section A.12 (online supplemental appendix).

LIMITATIONS

This study was conducted on one data source, MIMIC-III; hence, the validity of the observations must be confirmed using other data sources. Furthermore, an investigation of different care units may provide additional insights. The analysis of phenotype importance is produced on the RF, whose accuracy is superior to that of the baselines but not as good as that of the LSTM. This is limited by the poor computational efficiency of the SHAP values for the LSTM. Moreover, phenotypes annotated as transient are present only until a new clinical note appears in the timeline. This is inconvenient because phenotypes may be prematurely considered as not present because the next available clinical note did not mention them.

Acknowledgements We thank Dr Rick Sax, Dr Garima Gupta and Dr Matt Wiener for their feedback throughout this research; Dr Garima Gupta, Dr Deepa (MRSH) and Dr Ashok (MS) for helping us create gold-standard phenotype annotation data.

Contributors JZ, LDBT, AT and JI conceived the experiments and analysed the results. LDBT and AT conducted the experiments. JI, VB and YG reviewed the research and manuscript. All authors approved the manuscript. YG is the corresponding author and guarantor of this work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests This research is solely funded by Pangaea Data Limited. JZ, LDBT, AT, VG and YG are employed by Pangaea Data Limited. JI is employed by the Queen Mary University of London, but this study was conducted when JI was employed by Pangaea Data Limited. YG is also employed by Imperial College London and Hong Kong Baptist University, but YG supervised this study at Pangaea Data Limited.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The MIMIC-III dataset is publicly available at https://mimic.mit.edu/ (accessed in November 2022). The code, gold labels and other data cannot be shared publicly due to its proprietary nature.

ORCID iDs
Jingqing Zhang http://orcid.org/0000-0002-8822-736X
Luis Daniel Bolanos Trujillo http://orcid.org/0000-0003-0016-1288
Ashwani Tanwar http://orcid.org/0000-0001-9032-4348
Julia Ive http://orcid.org/0000-0002-3931-3392
Yike Guo http://orcid.org/0000-0002-3075-2161

REFERENCES

1 Nhs digital annual report and accounts 2019 to 2020, 2020. Available: https://www.gov.uk/government/publications/nhs-digital-annual-report-and-accounts-2019-to-2020
2 Johnson AEW, Ghassemi MM, Nemati S, *et al*. Machine learning and decision support in critical care. *Proc IEEE Inst Electr Electron Eng* 2016;104:444–66.
3 Kong H-J. Managing unstructured big data in healthcare system. *Healthc Inform Res* 2019;25:1–2 /pmc/articles/PMC6372467/https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC6372467/.
4 Harutyunyan H, Khachatrian H, Kale DC, *et al*. Multitask learning and benchmarking with clinical time series data. *Sci Data* 2019;6:1–18.
5 Subudhi S, Verma A, Patel AB, *et al*. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *NPJ Digit Med* 2021;4:87 http: //www.nature.com/articles/s41746-021-00456-x.
6 Alves T, Laender A, Veloso A. Dynamic Prediction of ICU Mortality Risk Using Domain Adaptation. In: *2018 IEEE International Conference on Big Data (Big Data*, 2018: 1328–36.
7 Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat* 2012;33:777–80.
8 Cooley-Rieders K, Zheng K. Physician documentation matters. using natural language processing to predict mortality in sepsis. *Intell Based Med* 2021;5:100028.
9 Liu C, Ta CN, Rogers JR, *et al*. Ensembles of natural language processing systems for portable phenotyping solutions. *J Biomed Inform* 2019;100:103318 http://www.sciencedirect.com/science/article/pii/S1532046419302370
10 Zhang J, Bolanos Trujillo L, Li T. Self-Supervised Detection of Contextual Synonyms in a Multi-Class Setting: Phenotype Annotation Use Case.. In: *Proc Conf Empir Methods Nat Lang Process*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics;, 2021: 8754–69. https://aclanthology.org/2021.emnlp-main.690
11 Alsentzer E, Murphy J, Boag W. Publicly Available Clinical {BERT} Embeddings.. In: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. . Minneapolis, Minnesota, USA: Associa- tion for Computational Linguistics;, 2019: 72–8. https://www.aclweb.org/anthology/W19-1909
12 Köhler S, Gargano M, Matentzoglu N, *et al*. The human phenotype ontology in 2021. *Nucleic Acids Res* 2021;49:D1207–17 https://pubmed.ncbi.nlm.nih.gov/33264411/
13 Lundberg SM, Lee SI, *et al*. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, eds. *Adv Neural Inf Process Syst. vol. 30. Curran Associates, Inc*, 2017. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
14 Johnson AEW, Pollard TJ, Shen L, *et al*. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:1–9 www.nature.com/sdata/
15 Vaswani A, Shazeer N, Parmar N. Attention is All you Need.. In: *Adv neural Inf process Syst 30*. In, 2017: 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf
16 Arbabi A, Adams DR, Fidler S, *et al*. Identifying clinical terms in medical text using Ontology-Guided machine learning. *JMIR Med Inform* 2019;7:e12596.
17 Breiman L. Random forests.. *Mach Learn.* 2001;10:5–32 https://link.springer.com/ article/
18 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80 http://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf
19 Lasko TA, Bhagwat JG, Zou KH, *et al*. The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* 2005;38:404–15.
20 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46 https://journals.sagepub.com/doi/
21 Pham-Gia T, Hung TL. The mean and median absolute deviations. *Math Comput Model* 2001;34:921–36.
22 Kramer AA. Are ICU length of stay predictions worthwhile? *Crit Care Med* 2017;45:379–80 https://journals.lww.com/ccmjournal/Fulltext/2017/02000/Are_ICU_Length_of_Stay_Predictions_Worthwhile.36.aspx
23 Forte JC, van der Horst ICC. Comorbidities and medical history essential for mortality prediction in critically ill patients. *Lancet Digit Health* 2019;1:e48–9 www.thelancet.com/digital-health
24 Nielsen AB, Thorsen-Meyer H-C, Belling K, *et al*. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish national patient registry and electronic patient records. *Lancet Digit Health* 2019;1:e78–89 https://pubmed.ncbi.nlm.nih.gov/33323232/
25 Knaus WA, Wagner DP, Draper EA, *et al*. The APACHE III prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619–36 https://pubmed.ncbi.nlm.nih.gov/1959406/
26 Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (saps II) based on a European/North American multicenter study. *JAMA* 1993;270:2957—2963 https: //pubmed.ncbi.nlm.nih.gov/8254858/.
27 Georgiou E, Hadjibalassi M, Lambrinou E. The Impact of Pain Assessment on Critically Ill Patients' Outcomes: A Systematic Review. *Biomed Res Int* 2015:2015 /pmc/articles/PMC4628961//pmc/articles/PMC4628961/?report=abstracthttps: //www.ncbi.nlm.nih.gov/pmc/articles/PMC4628961/.
28 Bastos PG, Sun X, Wagner DP, *et al*. Glasgow coma scale score in the evaluation of outcome in the intensive care unit: findings from the acute physiology and chronic health evaluation III study. *Crit Care Med* 1993;21::1459—1465 https://pubmed.ncbi.nlm.nih.gov/8403953/
29 Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950;78:1–3 https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml.