Digital Health and Health, Health-Service, and Equity Outcomes in the Eastern Mediterranean Region: A Systematic Review of the Evidence and Its Relevance to Population-Level Indicators

Abstract

Aims: This systematic review synthesizes evidence of digital health's impact on health outcomes, services, equity, and demographic indicators in the Eastern Mediterranean Region.

Methods: Following PRISMA 2020 guidelines, a systematic search was conducted in PubMed , Scopus, Web of Science and IEEE Xplore for English-language studies published from January 2010 to 12 June 2026. Eligible studies evaluated digital health technologies in WHO EMR countries and reported health, health-service, health-equity, demographic, or implementation-related outcomes. Reference screening and targeted searches of WHO and WHO Regional Office for the Eastern Mediterranean resources were performed. Due to heterogeneity in study designs, technologies, populations, and outcomes, findings were synthesized narratively.

Results: Among 7,894 identified records, 11 studies met the inclusion criteria. Evidence was organized into five themes: digital health for maternal and reproductive prediction and decision support; maternal, neonatal, and health-service outcomes; health inequalities; policy, social, cultural, and infrastructural determinants; and population-level demographic indicators. Digital health demonstrated potential improvements in clinical decision-making, healthcare access, and service delivery. However, equity outcomes depended on factors such as digital literacy, socioeconomic status, infrastructure availability, and contextual adaptation. Implementation success was influenced by governance, interoperability, workforce capacity, sustainable investment, and user engagement. No study directly assessed macro-level demographic outcomes like life expectancy or population aging.

Conclusions: Digital health and artificial intelligence in the Eastern Mediterranean Region have the potential to improve access to healthcare services, quality of care, and health-system performance. However, evidence regarding their impact on long-term demographic changes remains limited, highlighting the need for further research.

Keywords: Digital health; equity; Systematic review

PROSPERO registration: CRD420261447192

Introduction

The Eastern Mediterranean Region (EMR) is undergoing substantial demographic, epidemiological, and health-system transition, marked by changing fertility patterns, rising life expectancy, population ageing, and a growing burden of chronic and non-communicable disease (1-3). These shifts are increasing demand for accessible, sustained, and high-quality health services while placing further pressure on health systems that already vary considerably in financing, infrastructure, workforce capacity, and institutional readiness across the region (1, 2). The challenges are most pronounced in settings marked by socioeconomic disadvantage, humanitarian crises, displacement, and constrained access to care. Strengthening health-system capacity while preserving equitable access to care has therefore become an important regional priority (1).

Digital health has emerged as a central component of health-system transformation and progress toward universal health coverage. It spans a broad set of technologies, including electronic health records (EHRs), mobile health (mHealth), telemedicine and telehealth, digital public health systems, clinical decision-support systems, and artificial intelligence (AI). These technologies can support service delivery by improving access to information and care, enabling remote monitoring, informing clinical decision-making, and strengthening continuity of care and health-system coordination. In resource-constrained settings, digital health may also help offset geographical and service-access barriers by extending selected services beyond conventional facility-based models (4, 5).

The potential benefits of digital health, however, should not be equated with demonstrated population-level effects. Digital interventions typically act through intermediate pathways—access to services, service utilization, quality of care, clinical decision-making, disease management, and continuity of care—and improvements along these pathways do not, on their own, establish effects on population-level indicators such as life expectancy, healthy life expectancy, population ageing, or overall mortality. This distinction is particularly important when interpreting evidence drawn from heterogeneous sources, including predictive AI models, observational studies, implementation research, and evaluations of health-service interventions (6, 7).

Nor is the distribution of digital health's benefits likely to be uniform across populations. Differences in digital literacy, affordability, connectivity, infrastructure, health-system capacity, and access to technology shape who can use and benefit from digital health services. Vulnerable groups—including older adults, displaced and refugee populations, and socioeconomically disadvantaged communities—may face greater barriers to digital access. Digital health may therefore widen access and improve service delivery for underserved populations in some contexts, while insufficient attention to structural and contextual barriers may reproduce or deepen existing inequalities in others. Whether digital health contributes to equitable outcomes depends not only on the technology itself but also on the context in which it is designed, implemented, and integrated into the health system (1, 5).

Evidence from the EMR illustrates this contextual complexity. Regional assessments have documented marked differences among countries in digital-health maturity, infrastructure, governance, workforce capacity, financing, interoperability, and implementation readiness (1, 2). Cultural and linguistic characteristics, and the extent to which technologies are adapted to local needs, further shape their acceptability and uptake; studies conducted in the region have emphasized the importance of user engagement, contextual adaptation, appropriate information content, and alignment between digital solutions and local clinical and social needs (8-10).

These considerations are especially relevant in a region characterized by marked socioeconomic and health-system heterogeneity. Despite the rapid expansion of digital health initiatives, the evidence linking digital health technologies to health, service-use, demographic, and equity-related outcomes in the EMR remains fragmented, drawn from studies that differ substantially in design, population, intervention, outcome definition, and level of analysis. Consequently, it remains unclear what types of health and service-related outcomes have actually been evaluated, whether evidence exists for broader population-level outcomes, and which contextual factors influence the potential benefits and equitable implementation of digital health across the region.

Accordingly, this systematic review aimed to synthesize the available evidence on the association between digital health and health, health-service, and health-equity outcomes in the Eastern Mediterranean Region; identify key contextual factors influencing its implementation, accessibility, and benefits; and assess the extent of evidence addressing population-level demographic indicators.

Information & Methods

This study systematically reviewed the available evidence on the association between digital health and health, health-service, and health-equity outcomes in countries of the WHO Eastern Mediterranean Region (EMRO), while also identifying key contextual factors affecting its implementation, accessibility, and benefits and assessing the extent to which the available evidence addresses population-level demographic indicators . The design, conduct, and reporting of this study followed the PRISMA 2020 guidelines to ensure transparency, comprehensiveness, and reproducibility at every stage of the research (11). The study was registered with PROSPERO (registration number CRD420261447192).

This systematic review constitutes the first stage of a multi-stage study with a comparative-policy approach, intended to extract scientific evidence, valid indicators, and contextual factors relating digital health to demographic indicators, in order to inform the design of a second-stage quantitative study in Iran. Alongside primary research studies, official World Health Organization (WHO) policy documents, reports, and EMRO regional reports that met the eligibility criteria were also reviewed, since such documents articulate governance frameworks, policy priorities, and standardized digital health indicators relevant to identifying indicators that could be localized.

Research Questions

The review was guided by the following primary and secondary research questions:

Primary research question:

What evidence is available on the relationship between digital health technologies and health, health-service, and health-equity outcomes in countries of the WHO Eastern Mediterranean Region, and what evidence, if any, directly addresses population-level demographic indicators?

Secondary research question:

What social, cultural, infrastructural, health-system, governance, and policy factors influence the implementation, accessibility, equitable use, and potential benefits of digital health technologies in countries of the WHO Eastern Mediterranean Region?

The primary question was structured using the Population–Exposure–Outcome (PEO) framework (Table 1). The population comprised individuals and populations in countries of the WHO Eastern Mediterranean Region; the exposure was the use, implementation, or application of digital health technologies; and the outcomes encompassed health, demographic, health-service utilization, quality-of-care, and health-equity indicators. Contextual factors were treated separately, as potential implementation determinants and effect modifiers rather than as elements of the primary PEO framework.

Because this review forms the first stage of a broader programme of work intended to identify indicators for a subsequent quantitative study, the PEO framework deliberately specified macro-level demographic outcomes (e.g., population ageing, healthy life expectancy) alongside clinical, service-use, and equity-related outcomes. It was expected, on conceptual grounds, that few empirical studies would report directly on macro-level demographic indicators, since these typically require long-term, population-representative surveillance rather than study-level evaluation. Accordingly, this review captured intermediate clinical, service-use, and equity-related outcomes through which digital health may plausibly influence population health over time, and treated the presence or absence of direct empirical evidence on macro-level demographic indicators as a substantive review finding in its own right, reported explicitly in the Results.

Search Strategy

We systematically searched PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore for studies published from January 2010 to June 2026. Our search strategy focused on three core concepts: digital health technologies (e.g., eHealth, mHealth, AI), outcomes (e.g., clinical, health service, equity, population), and the WHO Eastern Mediterranean Region. We also considered contextual factors like policy, socioeconomic, and governance elements, but these were not primary search terms to maximize the inclusion of relevant studies on health outcomes, services, and equity.

The geographic source of participants or datasets in the included studies was identified using the names of the countries and territories comprising the WHO Eastern Mediterranean Region, together with region-specific geographic terms. General geographic terms such as “country” or “countries” were not used as substitutes for individual country names, as they lack the specificity required to precisely and reproducibly identify the study setting or data source. Geographic search terms were tailored to the indexing conventions and search capabilities of each database. Comprehensive, database-specific search strategies, including controlled vocabularies, keywords, Boolean operators, truncation symbols, and country-specific terms, are presented in Table 2. The search strategy was developed in accordance with the principles outlined in the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis. This guidance emphasizes that the selection of information sources should be based on the review question, the topic area, and the types of evidence expected, rather than on a predetermined or fixed number of databases (12).

The search was restricted to studies published in English between January 2010 and June 2026. The year 2010 was selected as the start date to encompass the period marked by the rapid development and adoption of novel digital health technologies. During the database search phase, no methodological or specific outcome-related restrictions were applied; this ensured the identification of studies addressing clinical outcomes, health services, health equity, and macro-level population outcomes, as well as the implementation and contextual aspects of digital health.

To identify potentially eligible studies not captured by the electronic database searches, the reference lists of all included studies and relevant systematic reviews were manually screened. Additionally, a supplementary manual search was conducted on the official websites and institutional repositories of the World Health Organization (WHO) and its Eastern Mediterranean Regional Office (EMRO) to identify relevant policy documents, regional reports, and other grey literature. Furthermore, the reference lists of relevant systematic reviews were examined to identify additional potentially eligible sources.

Eligibility Criteria

Inclusion Criteria

Studies were included if they met all of the following criteria:

1. Published between 1 January 2010 and 12 June 2026.
2. Geographic scope: Conducted, implemented, or specifically relevant to one or more countries of the WHO Eastern Mediterranean Region (WHO EMRO), or addressing the EMRO region as a whole.
3. Digital health exposure: Evaluated, described, implemented, or reported on at least one digital health technology, intervention, system, or application, including but not limited to artificial intelligence, machine learning, mHealth, telehealth, telemedicine, electronic health records, health information systems, clinical decision-support systems, and related digital health technologies.
4. Studies that examined at least one of the following: health or clinical outcomes; maternal, neonatal, fetal, or reproductive health outcomes; health-service utilization or quality of care; outcomes related to health equity, access to healthcare, or health inequalities; population-level demographic indicators, such as life expectancy, healthy life expectancy, or population ageing; or contextual factors relevant to the implementation, accessibility, equitable use, governance, sustainability, or potential benefits of digital health, including social, cultural, infrastructural, health-system, socioeconomic, and policy-related factors.
5. Conceptual relevance: Examined or provided evidence relevant to the relationship between digital health and health, health-service, health-equity, or population-level demographic outcomes, or provided relevant evidence on the contextual factors shaping the implementation, accessibility, or equitable benefits of digital health.
6. Evidence type: In addition to primary empirical studies identified through structured searches of PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore, relevant policy and grey-literature sources were eligible when they provided substantive evidence on digital health governance, policy, implementation, infrastructure, health equity, digital maturity, or related regional priorities. Relevant WHO and WHO/EMRO documents were identified through hand-searching official websites and institutional repositories and by screening the reference lists of relevant studies and systematic reviews.
7. Published in English.
8. Available in full text.

Exclusion Criteria

Studies were excluded if they met any of the following criteria:

1. Focused solely on the technical aspects of digital technologies without reporting a health or demographic outcome.
2. Were engineering or computer-science studies evaluating only artificial intelligence algorithms or computational models without assessing their application to population health or health care.
3. Were published as editorials, letters, commentaries, protocols, conference abstracts, dissertations, or theses, unless they constituted eligible official policy or grey-literature documents.
4. Were not available in full text.
5. Were duplicate publications.
6. Did not achieve an acceptable level of methodological quality according to the predefined quality-appraisal tools.
7. Were not directly aligned with the PEO framework or lacked direct conceptual relevance to the review question.

Study Selection

All records retrieved from the electronic databases were imported into EndNote reference-management software, and duplicate records were removed before screening. Study selection was conducted in two stages. In the first stage, the titles and abstracts of all retrieved studies were independently screened by two reviewers according to the predefined inclusion and exclusion criteria. In the second stage, the full texts of potentially eligible studies were independently assessed by the same two reviewers for eligibility. Disagreements were resolved through discussion, and a third reviewer was consulted whenever consensus could not be reached.

In addition to the electronic database search, the reference lists of all included studies and relevant systematic reviews were screened manually to identify any additional eligible studies that might not have been retrieved through the database searches. The study selection process is illustrated in the PRISMA 2020 flow diagram (Figure 1). Supplementary searches of grey literature were conducted in relevant organizational websites, institutional repositories, and WHO/EMRO resources. Search dates, sources, keywords, and navigation procedures were documented. Grey-literature documents were screened using the same predefined eligibility criteria applied to published studies. Eligible grey-literature sources were considered complementary contextual evidence and were not interpreted as equivalent to peer-reviewed empirical studies.

Data Extraction

Data were extracted independently by two reviewers using a standardized data-extraction form. Discrepancies between reviewers were resolved through discussion or, when necessary, by consultation with a third reviewer. For each included study, the following information was extracted: first author, year of publication, country, study objective, sample size, study design, measurement methods, and key findings (Table 3).

Quality Assessment

To strengthen the methodological rigor of this review and to account for the heterogeneity of the included study designs, the methodological quality of each study was assessed using a validated appraisal tool appropriate to its design. Because this review included qualitative studies, observational studies, technology-development studies, mixed-methods studies, systematic reviews, and policy documents related to digital health, no single quality-appraisal instrument was suitable for evaluating all study types; consistent with the recommendations of the PRISMA 2020 Statement, different validated critical-appraisal tools were used according to the design of each included study (11).

Quality assessment was conducted independently by two reviewers. Disagreements were resolved through discussion; when consensus could not be reached, a third reviewer made the final decision. The results of the quality appraisal were summarized in tabular form and considered when interpreting the findings and assessing the overall certainty of the evidence. Studies with serious methodological limitations, or that failed to meet the essential quality criteria of the relevant appraisal tools, were excluded from the final evidence synthesis.

Quality Assessment of Qualitative Studies

The methodological quality and risk of bias of the included sources were assessed independently by two reviewers using appraisal approaches appropriate to the study design. Because the included evidence comprised quantitative empirical studies, qualitative studies, and policy/grey-literature sources, different validated tools were used rather than a single instrument. For studies assessed with the Mixed Methods Appraisal Tool (MMAT), appraisal was conducted at the item/domain level using “Yes,” “No,” or “Can’t tell” responses, and the findings were not converted into an overall numerical or categorical quality score (13) .

Qualitative studies were assessed using the CASP Qualitative Checklist, with relevant domains reported descriptively and without assigning an overall quality category (14). The observational study was assessed using the Newcastle–Ottawa Scale (NOS), with the applicable domains and star-based assessment reported without assigning a separate overall categorical quality label (15). The AI prediction-model study by Sadr et al. was additionally assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST), with assessment of risk of bias across the participants, predictors, outcome, and analysis domains and consideration of applicability concerns (16). Policy and grey-literature documents were assessed using the AACODS checklist across Authority, Accuracy, Coverage, Objectivity, Date, and Significance and were interpreted as contextual/policy

evidence rather than primary empirical effectiveness evidence (17). The appraisal findings were used to inform interpretation of individual studies and the overall evidence base, particularly with respect to methodological limitations and risk of bias, rather than being collapsed into a single quality score across studies.

Quality Assessment of Policy Documents and Grey Literature

Because this systematic review included not only empirical studies but also policy documents, official WHO reports, and regional action plans, the quality and potential risk of bias of these documents were assessed using the AACODS (Authority, Accuracy, Coverage, Objectivity, Date, and Significance) checklist, a validated appraisal tool developed specifically for evaluating the credibility and methodological quality of grey literature and other non-peer-reviewed sources included in systematic reviews (17).

The appraisal was conducted across six domains:

- Authority: the credibility of the issuing organization (e.g., the WHO Regional Office for the Eastern Mediterranean or WHO headquarters), the expertise of the authors, and the authenticity of the source.
- Accuracy: whether the evidence and recommendations were supported by clearly described methodologies, such as regional surveys or health information systems, and whether appropriate references were provided.
- Coverage: the scope, completeness, and relevance of the document in addressing health-system policy questions and the demographic indicators of interest.
- Objectivity: the presence of any organizational, political, or commercial bias that could influence the recommendations or digital health action plans.
- Date: the timeliness of the document and its relevance to demographic transition and advances in digital health during the review period (2010–2026).
- Significance: the direct relevance of the document to the objectives of the review and its potential applicability to the Iranian health system.

The quality assessment of policy documents and grey literature was performed independently by two reviewers, and disagreements were resolved by consultation with a third reviewer. All policy documents included in this review met the predefined AACODS quality criteria (18).

Data Synthesis

Given the substantial heterogeneity among the included studies with respect to study design (observational, qualitative, mixed-methods, interventional, technology-development, systematic review, and policy documents), the diversity of digital health technologies evaluated (e.g., electronic health records, telemedicine, mobile health, artificial intelligence, clinical decision support systems, wearable technologies, and electronic registry systems), the wide range of demographic outcomes considered (including fertility, population ageing, healthy life expectancy, health equity, mortality, and health inequalities), and the variation in outcome-measurement methods—a meta-analysis was not considered appropriate. Findings were therefore synthesized using a narrative synthesis approach, following the framework proposed by Popay et al. Consistent with Popay et al.'s guidance on narrative synthesis, the synthesis process was conducted in four interconnected stages: (1) developing a preliminary theoretical model of how digital health interventions were expected to influence outcomes, conceptualized as a pathway from intervention → intermediate mechanism (e.g., improved access to information, enhanced clinical decision support, and increased service utilization) → clinical/service-level outcome → a hypothesized but not directly observed population-level pathway; (2) conducting a preliminary synthesis of the findings within this model; (3) examining relationships within and across studies, including the potential for data overlap and the historically controlled designs identified below; and (4) qualitatively assessing the robustness of the synthesis in light of the heterogeneity of the included study designs and the limited number of studies within each outcome domain. (19).

Titles and abstracts were screened according to the predefined eligibility criteria, and the full texts of eligible studies were subsequently reviewed for data extraction.

Findings

Study Identification and Selection

The study selection process is illustrated in the PRISMA 2020 flow diagram (Figure 1). A total of 7,894 records were identified through systematic searches of four electronic databases: PubMed (n = 551), Scopus (n = 6,158), Web of Science (n = 1,126), and IEEE Xplore (n = 59). After removal of 99 duplicate records, 7,795 studies remained for title and abstract screening. At this stage, 6,558 records were excluded, including 5,126 records with irrelevant titles, 1,379 studies conducted outside the WHO Eastern Mediterranean Region (WHO EMRO), and 53 conference papers, leaving 1,237 articles for full-text assessment.

Following full-text assessment, 1,226 articles were excluded: 1,207 did not meet the predefined eligibility criteria, and 19 did not achieve an acceptable level of methodological quality based on the relevant critical-appraisal tools. Ultimately, 11 studies met all eligibility criteria and were included in the systematic review and final synthesis.

Given the substantial heterogeneity in study design and the diversity of outcome-measurement methods, meta-analysis was not feasible. Findings were therefore categorized and reported using a synthesis approach based on the framework proposed by Popay et al., under four predetermined analytical themes: (1) digital health and fetal, maternal, and reproductive prediction/decision-support outcomes; (2) digital health and maternal, neonatal, and health-service outcomes; (3) digital health and health inequalities; and (4) policy, cultural, social, and infrastructural factors. A fifth subsection then reports explicitly on the extent of direct evidence for macro-level demographic indicators.

Characteristics of the Included Evidence

The study selection process is illustrated in the PRISMA 2020 flow diagram (Figure 1). A total of 7,894 records were identified through systematic searches of four electronic databases: PubMed (n = 551), Scopus (n = 6,158), Web of Science (n = 1,126), and IEEE Xplore (n = 59). After removal of 99 duplicate records, 7,795 studies remained for title and abstract screening. At this stage, 6,558 records were excluded, including 5,126 records with irrelevant titles, 1,379 studies conducted outside the WHO Eastern Mediterranean Region (WHO EMRO), and 53 conference papers, leaving 1,237 articles for full-text assessment.

Following full-text assessment, 1,226 articles were excluded: 1,207 did not meet the predefined eligibility criteria, and 19 did not achieve an acceptable level of methodological quality based on the relevant critical-appraisal tools. Ultimately, 11 studies met all eligibility criteria and were included in the systematic review and final synthesis.

1. Digital Health and Fetal, Maternal, and Reproductive Prediction/Decision-Support Outcomes

Sadr et al. (2025) developed a hybrid artificial intelligence model combining CNN-GRU with KNN, MLP, and LightGBM to predict clinical pregnancy success following IVF. The model was developed and evaluated using 500 medical records of couples undergoing IVF treatment. The proposed hybrid model demonstrated superior performance compared with all baseline models, achieving an accuracy of 94.67%, precision of 94.12%, sensitivity of 92.12%, an F1 score of 92.34%, and an AUC-ROC of 0.96, compared with 91.23% for KNN, 91.02% for LightGBM, 91.02% for CNN, and 91.02% for GRU. A paired t-test also demonstrated a statistically significant improvement in model performance ($p < 0.001$). The most important predictors identified by the model were female age, AMH level, and embryo quality. These findings are consistent with well-established clinical determinants of fertility and support the clinical relevance of the model, particularly in terms of the interpretability of its input features (20).

Farzandipour et al. (2017) adopted a fundamentally different approach, focusing on the design of an information infrastructure rather than a predictive algorithm. Using a Delphi methodology and two rounds of consultation with 50 infertility specialists, they identified the information requirements for electronic health records (EHRs) in Iranian infertility centers. Of the 261 initially proposed information items, 223 were endorsed by expert consensus, comprising 204 clinical items and 57 paraclinical items. The endorsed paraclinical items included hormonal tests (LH, FSH, AMH, progesterone, estrogen, prolactin, and testosterone), genetic tests (karyotyping and AZF), imaging

procedures (HSG, hysteroscopy, and ultrasonography), and semen analysis. The endorsed clinical items included menstrual history; pregnancy history (live birth, miscarriage, and stillbirth); medical history (including endometriosis, diabetes, and hypertension); previous infertility treatments (IVF, ICSI, and IUI); and a family history of infertility and recurrent miscarriage (10).

2. Digital Health and Maternal, Neonatal, and Health-Service Outcomes

Saleh et al. (2025) reported that an AI-, health-, and gamification-based intervention was associated with increased uptake of prenatal care, ultrasound examinations, and diagnostic testing, increased supplement use and full-term delivery, and reduced neonatal complications, in a social-experiment design with a historical (pre-intervention) comparison group (21); because the comparison was historical rather than concurrent and randomized, these associations may be affected by secular trends in care delivery and should be interpreted with corresponding caution. Venkateswaran et al. (2022), in the only cluster-randomised controlled trial included in this review, found that combining electronic health records with a clinical decision-support system improved screening and management of anaemia, hypertension, and gestational diabetes, although no significant difference was observed in overall adverse pregnancy outcomes between trial arms (22).

3. Digital Health and Health Inequalities

study by Saleh et al. (2025) reported that AI-enabled mobile health (mHealth) interventions among Syrian refugee women and socioeconomically disadvantaged women in Lebanon were associated with improved access to antenatal care, increased uptake of recommended prenatal screening, and reduced several adverse neonatal outcomes (21) . By contrast, AlArab et al. (2023) found that, despite the availability of electronic health records (EHRs), Syrian refugee women in Lebanon continued to experience higher rates of caesarean delivery, spontaneous abortion, shorter interpregnancy intervals, and lower breastfeeding rates after displacement, underscoring the influence of social and economic determinants that lie beyond the reach of technology alone (23).

Substantial disparities in internet access ranging from 18% to 100% across different settings were identified as a key factor contributing to inequities in the utilization and benefits of digital health services (2). At the levels of design and adoption, Moulaei et al. (2021) developed a COVID-19 self-care application for Iranian pregnant women. In the first phase, information needs were assessed among 15 obstetricians and 15 pregnant women, followed by an evaluation involving 36 pregnant women (11 with COVID-19 and 25 without COVID-19) in the second phase. Of the 41 identified information needs, 35 were endorsed and incorporated into the application, with its features organized into four domains: user profile, lifestyle, disease management and control, and app functionalities. The overall usability score was 7.94 out of 9, indicating good usability, with the highest satisfaction reported for app features (8.62) and the lowest for learnability (7.49) (9).

Pouriayevali et al. (2023), in a qualitative study based on semi-structured interviews with 14 Iranian pregnant women, identified seven major themes representing practical needs for the design of a pregnancy self-care application: nutrition; education, including warning signs and birth preparation; fetal health monitoring, including fetal movement counting and heart rate; relaxation and mental health; general health; risk factors; and physical activity. A key finding was that although women expressed a willingness to use such an application, the limited comprehensiveness of existing applications led them to rely on general internet searches to address some of their informational needs. This finding provides direct evidence that inequities in the benefits derived from digital technologies are not attributable solely to lack of access, but also to the quality, comprehensiveness, and contextual relevance of locally adapted digital health content (8).

4. Direct Evidence for Population-Level Demographic Indicators

None of the eleven studies or documents included in this review directly measured a macro-level demographic indicator, such as life expectancy, healthy life expectancy, or the total fertility rate, at the national level. Nevertheless, three studies partially and indirectly addressed indicators that are commonly reported at the population level in the reproductive health and demographic literature: age at first pregnancy, interpregnancy interval, and fetal or neonatal mortality rates per 1,000 births. A closer examination of these measures illustrates why none can substitute for genuine population-level surveillance evidence.

AlArab et al. (2023) was the only study in this review to directly report two conventional demographic indicators of fertility: age at first pregnancy (mean, 21.6 years, with 21.4% of pregnancies occurring before age 18 in the Lebanese-only group) and interpregnancy interval (decreasing from 2.32 to 1.76 years). These are precisely the types of indicators routinely reported

in standardized national demographic and health surveys. However, three fundamental limitations distinguish these data from genuine population-level evidence. First, the study population was restricted to 1,065 registered Syrian refugee women captured in a specific EHR database (Sijilli), rather than a representative sample of the general population. Second, the study used a descriptive secondary-analysis design rather than a continuous population-surveillance system. Most importantly, these indicators measured the consequences of displacement rather than the effects of a digital health intervention. No causal relationship or even a direct associationbetween digital technology use and these outcomes was tested. In this context, the EHR served solely as a retrospective data-recording tool rather than as the intervention variable under investigation (23).

Venkateswaran et al. (2022), which employed the most rigorous design in this review—a cluster-randomized controlled trial—reported fetal mortality using a commonly used population-level metric: 6 per 1,000 in the control group versus 7 per 1,000 in the intervention group, a difference that was not statistically significant. Unlike the AlArab study, these figures were derived from an actual intervention comparison. However, two important limitations prevent their interpretation as a population-level indicator. First, the study population was limited to 6,367 pregnant women attending participating clinics rather than a nationally representative sample. Second, the lack of statistical significance and the fact that the direction of the difference was contrary to the expected direction means that the study provides no evidence of improvement in a population-level indicator and cannot establish a beneficial effect of the digital intervention on this specific outcome (22).
Mominkhan et al. (2025), the only study in this review primarily focused on mortality, demonstrated a similar pattern on a larger scale, including a total of 13,532 live births. The study reported a 10.7% numerical reduction in the neonatal mortality rate, from 10.3 to 9.2 per 1,000 live births; however, this difference was not statistically significant ($p = 0.856$), and the study employed a retrospective design without a concurrent control group (24).
Taken together, these three studies, each addressing a different dimension of the demographic-indicator continuum fertility and interpregnancy spacing, fetal mortality, and neonatal mortality—form a three-part but methodologically heterogeneous pattern. None fully meets the requirements for valid population-level evidence: a representative population sample combined with a statistically significant causal relationship between a digital intervention and the demographic outcome within a continuous surveillance framework. Each study falls short on at least one of these essential criteria. AlArab et al. (2023) provides no tested causal relationship with a digital intervention; Venkateswaran et al. (2022) and Mominkhan et al. (2025) evaluated intervention-related associations but did not demonstrate statistically significant effects; and all three studies remained restricted to populations within specific contextual settings (22-24).

5. Contextual, Policy, and Governance Evidence (Secondary Evidence Stream): Policy, Cultural, Social, and Infrastructural Factors

The WHO Global Strategy on Digital Health (2021) emphasized the need for comprehensive national digital health policies, expanded digital infrastructure, standardization and interoperability of health information systems, data protection, workforce capacity-building, and sustainable investment as fundamental prerequisites for the successful implementation of digital health (25).
The WHO regional report (2024) indicated that, although many EMR countries have established national digital health policies, significant challenges remain, including insufficient funding, shortages of skilled human resources, inadequate infrastructure, and the absence of common standards. The report also identified Iran as one of the regional leaders in digital health strategy, data protection, and the development of digital health applications. Furthermore, the WHO Regional Action Plan for Digital Health 2024–2028 (published 2025) provided an operational framework for advancing digital health across the region's 22 countries. Developed through an evidence review, a regional survey, and semi-structured interviews with 19 regional experts, the action plan classified countries into three levels of digital-health maturity—emerging, developing, and advanced—and proposed tailored SMART objectives and context-specific actions for each level. The framework identified four strategic priorities: strengthening governance and enabling environments, supporting people-centred health systems, enhancing national capacity, and fostering multisectoral partnerships for innovation and research. It also introduced monitoring and evaluation indicators, digital-maturity metrics, sustainability and scalability strategies, and specific recommendations for countries affected by emergencies. These findings underscore that the successful implementation of

digital health depends not only on technological advancement but also on effective governance, sound policymaking, sustained investment, and cross-sectoral collaboration (26).

At the national level, studies by Farzandipour et al. (2017), Moulaei et al. (2021), and Pouriayevali et al. (2023) similarly demonstrated that the successful design and implementation of digital health systems requires active user engagement, adaptation to local cultural and clinical contexts, and the development of context-specific, locally tailored solutions (8-10).

Discussion

Principal Findings

This review found that digital health technologies in the Eastern Mediterranean Region (EMR) are associated with improvements in selected clinical, health-service, and equity-related outcomes—particularly antenatal care utilization, aspects of care quality, clinical decision-making, and some maternal and neonatal outcomes—but that no included study provided direct empirical evidence linking digital health to macro-level demographic indicators such as population ageing, healthy life expectancy, or overall life expectancy. This is an important distinction: the review's own PEO framework and its parent research programme were oriented toward population-level indicators, yet the empirical literature currently available for the EMR speaks almost entirely to intermediate, clinical, and service-level outcomes. We consider this gap, rather than a specific effect estimate, to be one of the more informative findings of this review, since it indicates that claims linking digital health to population ageing or healthy life expectancy in this region are not yet empirically supported and should be treated as hypotheses for future research rather than established relationships.

Comparison with the Broader Literature

These findings are broadly consistent with the World Health Organization's global and regional digital health frameworks, which similarly frame digital health as a means of strengthening health systems rather than as a technology with directly demonstrated population-level effects. In 2025, the World Health Assembly (WHA) approved the extension of the Global Strategy on Digital Health 2020–2025 through 2027 and initiated development of its next phase for 2028–2033, underscoring the growing policy importance of digital health for equitable, resilient, and people-centred health systems, even as direct outcome evidence at the population level remains limited globally and not only in the EMR (27).

Interpretation by Outcome Domain

With respect to maternal, neonatal, and fetal-related outcomes, this review found that artificial-intelligence-based models for fetal-mortality prediction and maternal-health management showed high predictive accuracy in the studies evaluated, and that AI- and gamification-based interventions were associated with increased antenatal care uptake and selected improvements in neonatal outcomes. These findings should, however, be interpreted as evidence of model performance and short-term service-level association rather than demonstrated clinical effectiveness at scale, since only one included study (Venkateswaran et al.) used a randomized design; most other associations were derived from observational, retrospective, or historically controlled comparisons that are susceptible to secular trends and residual confounding(22).

With respect to maternal, neonatal, and fetal-related outcomes, this review found that artificial-intelligence-based models for fetal-mortality prediction and maternal-health management showed high predictive accuracy in the studies evaluated, and that AI- and gamification-based interventions were associated with increased antenatal care uptake and selected improvements in neonatal outcomes. These findings should, however, be interpreted as evidence of model performance and short-term service-level association rather than demonstrated clinical effectiveness at scale, since only one included study (Venkateswaran et al.) used a randomized design; most other associations were derived from observational, retrospective, or historically controlled comparisons that are susceptible to secular trends and residual confounding (22). Two of the AI prediction studies (Alzakari et al. and Innab et al.) reported cardiotocography data sets with identical sample compositions (2,126 records, comprising 1,655 normal, 295 suspicious, and 176 pathological cases), raising the possibility that they drew on overlapping or shared source data; because this could not be confirmed from the published reports, the two studies were treated as independent for narrative synthesis purposes, but this potential overlap should be considered when weighing the cumulative weight of AI-based evidence in this review (28, 29).

With respect to health inequalities, this review highlights the dual nature of digital health. In some studies, digital technologies were associated with improved access to antenatal care among underserved populations, such as Syrian refugee women in Lebanon. In others, broader

socioeconomic determinants appeared to outweigh the benefits of technology, coinciding with persistent disparities (30). These findings align with recent concerns that implementing digital health interventions without addressing structural and contextual determinants may reproduce, or even widen, existing health inequities. The WHO regional report on digital health in the Eastern Mediterranean Region (2024) similarly acknowledged that, despite the adoption of national digital health policies in many countries, the digital divide, inadequate funding, shortages of skilled personnel, and the lack of interoperable standards remain major barriers to achieving equity in digital health (26). Furthermore, a 2025 WHO report found that only 21% of Member States conduct comprehensive safety assessments before and after implementing digital health technologies, while only 22% have established systems for monitoring the safety of digital health products findings that reveal a substantial gap between technological innovation and the governance mechanisms required to ensure its safe implementation (31).

Cultural Adaptation and Localization

With respect to cultural adaptation and localization, the present review found that the success of digital health interventions depends on their alignment with the cultural, linguistic, and contextual needs of the target population. A qualitative study by Pouriayevali et al. reported that Iranian pregnant women frequently relied on general internet searches to meet information needs that existing pregnancy applications did not comprehensively address, highlighting a gap between technology design and users' cultural and social expectations (8). This finding is consistent with international evidence emphasizing the importance of cultural adaptation in digital health interventions; recent research suggests that cultural adaptation is an iterative and systematic process that requires substantial resources, sustained engagement of end users and stakeholders, multidisciplinary collaboration, and ongoing evaluation throughout development and implementation (32).

Implications for the Second-Stage Study in Iran

This review was conducted as the first stage of a broader, multi-stage research programme intended to inform a subsequent quantitative study of digital health and demographic indicators in Iran. Because the review's design was partly shaped by that downstream purpose—for example, in its inclusion of WHO policy documents and its broad PEO framework—its findings should be read as informing the selection of candidate indicators and contextual factors for the second-stage study, rather than as an independent test of whether digital health influences population-level indicators in Iran specifically. This distinction is made explicit here to avoid the appearance that evidence was assembled to justify a predetermined conclusion.

Strengths, Limitations, and Implications

This review has several strengths, including PRISMA 2020-consistent reporting, PROSPERO registration, dual independent screening and data extraction, and the use of design-matched critical appraisal tools. It also has a number of limitations that should inform how its findings are used.

First, the eligibility criteria were intentionally broad spanning observational, interventional, qualitative, mixed-methods, technology-development, systematic-review, and policy-document sources—in order to comprehensively map the evidence base for a subsequent quantitative study. This breadth precluded meta-analysis and produced a synthesis that functions more as a systematic evidence map than as a pooled effect-size analysis; readers should not interpret the narrative groupings in the Results as indicating consistent, quantifiable effects across studies. Second, most included empirical studies were observational, cross-sectional, or historically controlled, making them susceptible to selection bias and residual confounding; only one included study used a randomized design. Third, design-matched appraisal tools (MMAT, the Newcastle–Ottawa Scale) were applied to several artificial-intelligence prediction-model studies for which they were not originally designed, and no formal certainty-of-evidence framework (e.g., GRADE or GRADE-CERQual) was applied across outcome domains; both should be regarded as methodological limitations rather than as an assessment of overall evidence certainty. Fourth, Although the final version of this review included searches of PubMed, Scopus, Web of Science, and IEEE Xplore, the search was restricted to English-language publications. Consequently, studies published in other languages in Eastern Mediterranean Region countries which may include a substantial body of locally relevant research published in Arabic, Persian, French, or Urdu were not eligible for inclusion. This language restriction may have introduced language bias into the evidence base. Fifth, the threshold for excluding studies on methodological-quality grounds was applied using the general quality criteria of each design-matched appraisal tool rather than a single prespecified numerical cut-off,

which involves an element of reviewer judgement. Sixth, the identification of WHO and grey-literature policy documents was not conducted with the same systematic search strategy as the database search, so its coverage should not be regarded as exhaustive. Finally, the review focused primarily on EMR countries; although this focus reflects the region's relevance to the parent research programme in Iran, it may limit the generalizability of the findings to other settings, and the absence of a formal comparison against prior systematic reviews in this field means the novelty of this synthesis relative to earlier work has not been independently verified.

Based on these findings, several implications can be drawn, offered as considerations for policymakers and researchers rather than as directives supported by causal evidence from this review. National digital health strategies could be aligned with the WHO Global Strategy on Digital Health and the WHO Regional Action Plan for Digital Health 2024–2028, with attention to vulnerable populations, including refugees and residents of underserved communities. Expanding health insurance and reimbursement coverage for digital health infrastructure, user education, and technical support may help narrow the digital divide, although this review did not evaluate such policies directly. The design and implementation of digital health technologies would likely benefit from active end-user involvement and adaptation to cultural, linguistic, and religious context, given the qualitative evidence on unmet information needs identified in this review. Most importantly, future primary research should prioritize longitudinal and comparative designs capable of directly measuring population-level demographic indicators, since this review found no existing empirical basis for such claims in the EMR literature. This review was not designed to evaluate intervention effectiveness or generate policy recommendations; rather, it aimed to characterize and synthesize the available evidence on digital health and AI-related applications in the Eastern Mediterranean Region. Due to substantial heterogeneity in study designs, interventions, outcomes, and evidence sources, formal outcome-level certainty assessments such as GRADE/GRADE-CERQual were not undertaken. Future focused systematic reviews with comparable outcomes and feasible quantitative synthesis may apply these frameworks to support evidence-informed policy and guideline development.

Conclusion

This systematic review demonstrates that digital health technologies, including artificial intelligence (AI), mobile health (mHealth), telemedicine, electronic health records (EHRs), and clinical decision support systems (CDSSs), are primarily associated with improvements in healthcare access, service delivery, clinical processes, decision-making, and selected maternal, neonatal, and equity-related outcomes across countries of the Eastern Mediterranean Region (EMR). AI-based predictive models showed promising performance in specific clinical applications; however, these findings represent model-level accuracy and do not constitute evidence of demonstrated population-level health impact. Importantly, this review identified a major evidence gap: no included study directly evaluated the relationship between digital health implementation and population-level demographic outcomes, including life expectancy, healthy life expectancy, or population ageing. Therefore, potential pathways linking improved digital-health-enabled healthcare delivery to broader demographic and population health changes remain hypothetical and require dedicated evaluation. Future research should move beyond implementation and service-level outcomes by incorporating longitudinal, comparative, and population-based study designs. Such efforts, together with strengthened governance, infrastructure, workforce capacity, equity-oriented approaches, and context-specific implementation strategies, are needed to determine whether and under what conditions digital health can contribute to measurable improvements in population health.

Ethical Considerations

As this study was a systematic review based exclusively on previously published data, ethical approval and informed consent from participants were not required. Nevertheless, all stages of the review were conducted in accordance with the principles of research integrity, including accurate citation practices and adherence to established ethical standards for scientific research.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Author Contributions

Mahdieh Joukar: contributed to writing the Results, Discussion, and Conclusion sections; participated in the systematic literature search, study screening and selection, data extraction, and quality assessment of the included studies; contributed to the synthesis and analysis of the findings; prepared the initial manuscript draft; and critically revised the scientific content.