ÓBUDAI EGYETEM
OBUDA UNIVERSITY

Habilitációs Tézisfüzet

Dr. Zrubka Zsombor János

Advancing the Social Construction and Institutionalisation of Digital Health Innovations

Strengthening Definitions, Evidence Frameworks, and Reporting Standards for Market Access

Innováció
Menedzsment
Doktori Iskola

Budapest, 2025.

Table of Contents

Background

Digital health technologies (DHT) are defined as the “use of information technology and electronic communication tools within the delivery of healthcare services” (Canada Health Infoway 2021). DHTs, including artificial intelligence (AI) tools, software and devices used for research and development, diagnosis, prognosis, monitoring, and therapy (Kasoju et al. 2023). In this booklet, the term DHT will be used interchangeably with other terms such as digital health intervention (DHI), digital biomarker (DBM) based intervention or digital medical device.

Over the past decade, market uptake of DHTs has accelerated, and the COVID-19 pandemic further catalysed this growth (Golinelli et al. 2020). In 2022, the global DHT market was 211 bn USD, forecasted to grow at a compound annual rate of 18.6% until 2030 (Kasoju et al. 2023). In 2017, over 318 000 health apps were available in app stores, with 200 added each day (IQVIA Institute for Human Data Science 2017).

In parallel, large scale legal and regulatory frameworks were introduced in the United States (US) and the European Union (EU) to protect markets, population health and fundamental rights. These include the US Food and Drug Administration's (FDA) Medical Device Amendments of the 1976 Federal Food, Drug and Cosmetic Act, the EU Medical Device Regulation (MDR) the EU In Vitro Diagnostic Medical Devices Regulation (IVDR), and the EU Artificial Intelligence Act, the General Data Protection Regulation (GDPR) (ICLG Group 2024). However the regulatory landscape of DHTs remained scattered and lags behind technological development (ICLG Group 2024). The rapid diffusion of ChatGPT, reaching 100 million users within 2 months after its launch exemplifies this tension (Hu 2023).

Unlike the consolidated pharmaceutical industry (Shepherd 2018), the DHT market is fragmented, comprising of small innovative firms specialised in rapidly changing technologies (Kasoju et al. 2023). Fragmented markets and elevated regulatory standards increase development costs and the price for patients

(Khan et al. 2024; Heinemann 2021). To facilitate patient-access, many developers seek public funding. Yet, public financing and health technology assessment (HTA) processes (i.e., structured assessments to inform decision makers) represent another diverse landscape (Tarricone, Petracca, and Weller 2024).

Several innovation theories provide perspective on this scattered landscape. The Social Construction of Technology (SCOT) argues that technologies are not defined solely by their function or utility, but by the meanings attached to them by relevant social groups (Pinch and Bijker 1984; Bijker, Hughes, and Pinch 1993). Currently, digital health is in a phase of interpretive flexibility, when DHTs may signify the next big thing to investors, efficiency for health managers, empowerment for patients, or risk for regulators. Through the interactions of social groups, these contested meanings advance toward closure, when interpretations stabilise through standards, guidelines, or institutional practices. Structuration Theory conceptualizes how structures (e.g., institutions, rules, and resources) shape action while

simultaneously being shaped by social actors (Giddens 1984). Evidence frameworks or reporting standards, for example, both structure and shaped by the innovation practices of developers, policymakers, and clinicians. Evolutionary economics highlights path dependency, whereby innovations are embedded in pre-existing institutional and market arrangements (Nelson and Winter 1982; Dosi 1982). DHTs are initially fitted into structures developed for pharmaceuticals, such as evidence frameworks, reimbursement models, and HTA procedures. Over time, DHTs and institutional frameworks co-evolve until more stable market and institutional configurations emerge (Freeman 1987).

A substantial part of my research has focused on methodological developments for conducting and reporting HTA studies of DHTs. I have led research for the Digital Health Special Interest Group (DH-SIG) of ISPOR (the "leading professional society for health economics and outcomes research globally") and contributed to the TKP2020-NKA-02 and TKP2021-NKTA-36 projects (Assessment of Digital Health Technologies: Efficacy,

Safety and Societal Benefits) funded by the National Research, Innovation and Development Fund of Hungary.

Through a series of descriptive and prescriptive studies, my work has evaluated and shaped the methodological and reporting quality of evidence syntheses and health economic evaluations of DHTs. In this way, my research has contributed both to the discourse over the meaning of DHTs as well as the institutional frameworks shaping their innovation and market access. This thesis booklet addresses the following questions.

1. In the lack of standardized DHT taxonomies, how can DHTs be uniquely defined?
2. How COVID tracing apps balance between data privacy and public health interests?
3. Beyond the traditional HTA modules, what evidence is required by payers for funding decisions of DHTs?
4. What do structured assessments reveal about the quality of clinical evidence and reporting of various DHTs?
5. How to select a reporting guideline for a medical AI study?

Results

In the lack of standardized DHT taxonomies, how can DHTs be uniquely defined?

Publication 1: Scoping review of systematic reviews of digital biomarkers

Rationale

Digital biomarkers (DBM) are “objective, quantifiable, physiological and behavioural measures collected using digital devices that are portable, wearable, implantable or digestible” (Babrak et al. 2019). DBMs allow data collection at times or places that are not possible in clinical settings, supporting personalised therapy, monitoring or prognosis [1]. Systematic reviews usually formulate research questions using the PICO framework (Patient, Intervention, Comparator, Outcome). Given the absence of a standard DBM taxonomy, we examined to what extent the PICO elements of systematic reviews of DBMs can be mapped to existing international medical taxonomies: the WHO International Classifications of Diseases 11th revision (ICD 11) (WHO 2020a), the International Classification of Health Interventions

(ICHI) (WHO 2020b), and the International Classification of Functioning, Disability and Health (ICF) (WHO 2001).

Methods

We performed a systematic scoping review by searching systematic reviews of DBMs in PubMed and the Cochrane Library published in 2019-2020. PICO questions for identified DBM-based interventions were mapped as follows. patients to ICD 11, interventions to ICHI, comparators to ICHI, outcomes to ICF. Details are provided in [1].

Results

From 375 records, we identified 31 systematic reviews. Of the 31 studies, 25 (80.6%) mapped to ICD 11 disease categories, while 6 (19.4%) concerned healthy populations (e.g., employees, students or healthy individuals) not represented in ICD 11. A small number of interventions (2/31, 6.5%) could not be coded with the WHO ICHI tool.

New results

Most PICO questions of DBM systematic reviews could be coded using WHO tools, but interventions on healthy populations were not classifiable. As prevention and risk reduction in healthy populations are key applications of DHTs, this gap in international medical taxonomies should be addressed to capture all use cases of DHTs.

Publication 2: The Usefulness of Digital Health Terms for Outcomes Research

Rationale

The COVID pandemic accelerated the adoption of DHTs (Koonin et al. 2020). Expectations were high regarding improvements in care quality, healthcare analytics (Wake et al. 2020), consumer empowerment and informed decisions, or tackling health inequity or poor access to services (FDA 2020; WHO 2019). However, inconclusive evidence summaries revealed uncertain effects, methodological heterogeneity and a lack of consistent terminology related to DHTs (Stevenson et al. 2019).

Conducted by DH-SIG of ISPOR this research aimed to examine definitions of four umbrella terms: digital health,

eHealth, mHealth and telehealth / telemedicine; map their occurrence and evaluate their usefulness for outcomes research.

Methods

We conducted a systematic scoping review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline extension for Scoping Reviews (PRISMA-Scr) (Tricco et al. 2018). With a librarian, we combined filters for systematic reviews, Medical Subject Headings (MeSH) terms for digital health, and synonyms of "definition" to search PubMed/MEDLINE, Embase, Cochrane Library, and EconLit. English-language systematic reviews published in 2015-2020 were eligible, with a digital health-related MeSH term and a synonym of "definition" in their title or abstract. Screening, report identification, and data extraction were performed independently by 15, 10, and 7 pairs of reviewers, respectively. Differences were resolved by consensus. From full-text reports we extracted digital health-related terms and their definitions. Definitions were edited according to uniform cleaning rules, and categorised as

"original" (e.g., invented by the author), "adopted" (e.g., verbatim citations of a primary source), and "adapted" (e.g., existing definitions conceptually modified). Unique definitions were counted, and content analysis involved the observation of top 20 keywords via term frequency/inverse corpus frequency (TF/ICF), and word clouds [2].

Results

The search yielded 2610 records after deduplication, out of which 545 full text papers were examined for eligibility. Altogether, 236 systematic reviews contained a definition for a digital health-related term, out of which 134 defined one of the four umbrella terms [2].

We identified 4, 36, 50, and 52 unique definitions for digital health, eHealth, mHealth and telehealth / telemedicine, respectively. Between 2016 and 2019, in each year nearly 10 novel original definitions were observed. While the word "health" dominated most definitions, the text analyses revealed no characteristic words that differentiated the 4 umbrella terms [2].

Figure 1. Word clouds of digital health umbrella terms

Source: [2]

New results

Despite the proliferation of unique digital health-related definitions over time, the concepts were overlapping, and ambiguous about the functional domains of digital health. The DH-SIG recommended combining PICO (Population, Intervention, Comparator and Outcome) elements with specific secondary digital health terms (e.g., telesurgery, telerehabilitation). Relying on umbrella terms alone risks overly heterogenous evidence summaries of the studied interventions [2].

Publication 3: The PICOTS-ComTeC framework to define Digital Health Interventions

Rationale

Building on the DH-SIG scoping review, a qualitative content analysis mapped the information content of secondary digital health-related terms' definitions to 24 domains taken from common health research frameworks. Thes included PICOTS (Patients, Intervention, Comparator, Outcome, Timing, and Setting) (US Food and Drug Administration na), the Shannon-Weaver model of communication (Neuendorf 2017), the Agency for Healthcare Research and Quality quality domains (AHRQ na), technological features, geographic scope and the WHO's Classification of Digital Health Interventions (WHO 2018). In 101 included definitions, we identified 95 distinct information patterns [3].

Given this heterogeneity, the DH-SIG sought to develop a standardised minimum information framework to define DHTs.

Methods

We conducted an international 3-round Delphi consensus study with DH-SIG members. Volunteers with at least moderate familiarity with DHTs were included. Initial items were drafted by a core group, with additional items suggested by panel members during the first Delphi round. Consensus was then pursued in rounds two and three using predefined thresholds. The study followed the Conducting and Reporting Delphi Studies (CREDES) guideline (Junger et al. 2017) and the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Guidance for Developers of Health Research Reporting Guidelines (Moher et al. 2010). Further methodological details are provided in [3].

Results

Eighteen, 11, and 10 experts participated in the three Delphi rounds. The panel was diverse in geography, gender, age, expertise, and work setting. With high overall agreement, 9 domains (Patients, Intervention, Comparator, Outcome, Timing, Setting, Communication, Technology, Context) and 32 subcategories were retained.

The PICOTS-ComTeC framework was pilot tested on four breast cancer disease-management apps.

Table 1. The PICOTS-ComTec framework

Item	Explanation
Population Domain	Characterization of patients/ population(s)
Target Population/ Diagnosis	Diagnosis/ condition/ population (may be more than one)
Demographic Characteristics	Socio-demographics of population (e.g., age, gender, education)
Special User Characteristics	DHI relevant user characteristics (e.g., digital literacy, PC access)
Intervention Domain	Description of DHI Intervention including key components and interactions
Key Function/ Intended Use	Intended function (e.g., online screening to identify high-risk patients)
Modality	Design elements to achieve key function (e.g., behavioural, communication)
Limits of intervention	To specify those situations or thresholds where the DHI can be used, and beyond which the DHI should be replaced by face-to-face care
Comparator Domain	Non-DHI(s) or alternative DHI(s) with same function
Model of Care	Current model of care and/or clinical pathway, may be redesigned by DHI
Alternative Digital Health Interventions	DHI(s) with the same purpose (e.g., smart phone vs PC retinal screening)
Usual Care Alternatives	Usual treatment or care (e.g., compare with paper-based surveillance)
Outcomes Domain	Outcomes relevant to patients and other stakeholders
Health Benefits	Clinical and patient reported outcomes
Improved Care Structure or Process	Health care system improvements (e.g., access to care, adherence to guidelines, patient health literacy, self-management)
Social/ Societal Benefits	Humanistic, social, or societal effects (e.g., DHI could improve social support, or reduce stigma of a condition)
Safety	May reduce health related risks or improve patient safety
Non-health Related Risks	Non-health related risks including data privacy (e.g., unauthorized access and use of personal data)
Efficacy, Convenience, and Economic Benefits	DHIs could deliver the same outcome with greater efficiency, or less effort
Timing Domain	Timing and duration of treatment and follow-up
Timeliness	Timely delivery of services could improve outcomes (e.g., telestroke DHI to shorten time to thrombolysis could improve survival)
Frequency and Duration of Intervention	Increased DHI use may improve outcomes (e.g., increased use in cardiac rehabilitation associated with greater weight-loss)
Setting Domain	DHIs may increase access to or improve quality of health care. Potential benefits may vary by setting.
Care Setting	Settings where DHI may be useful include pre- and post-hospitalization, emergency care, primary and community care.
Patient Location	DHIs can bring care to the patient's location (e.g., in-home hospital care during COVID-19, public kiosks providing access to nurses).
Geographic Scope	DHIs can improve access to health care (e.g., rural Alaska). Culture may limit use (e.g., telehealth differences in Brazil vs Canada).

Communication Domain	DHIs may have different users with different roles. Function impacts frequency of interaction (e.g., post-surgical vs routine monitoring).
User	DHI users may vary (e.g., activity monitoring for patient lifestyle modification involving healthcare providers, or support groups).
Message	Unit of information collected and communicated by DHI (e.g., text, diagnostic image, or machine-readable data) impacts function.
Interaction Pattern	Differences in interactions (e.g., synchronous (real-time) or asynchronous) could impact outcomes in critical situations.
User Experience	Improving user experience may improve outcomes (e.g., when human factors were considered in digital interface design).
Technology Domain	Use of different technologies (i.e., communication channel, device, software, or system) may impact DHI performance.
Channel/ Medium	Channel selection may impact patient access and DHI effectiveness (e.g., DHIs that exclude patients without telephone access).
Device	DHIs involve devices or user interfaces that may vary in cost and accessibility (e.g., patient access to mobile phone vs PC)
Software	Algorithms (e.g., for machine learning) and software components (e.g., for security) used by DHIs may affect performance.
System	Compatibility with data standards (e.g., FHIR) and interoperability with larger healthcare systems may affect DHI potential.
Data Management	Considerations include data quality, timeliness, interoperability (e.g., with EHR), security, patient privacy, and legal requirements.
Context domain	Capture additional information that may influence the usability, access, or overall value of DHIs.
Regulatory status	The relevant regulatory category and authorization status for the DHI to identify appropriate comparators. (E.g., FDA approved or investigational)
Medical / legal liability	Specify if certain legal provisions influence the availability or effect of the . (e.g., can a medical expert give advice or only tests results can be communicated)
Financing	Specify if certain reimbursement or financing rules or pricing schemes influence the availability of functionality of the DHI. (e.g., in-app purchases, free from health service provider, subscription fee etc.)

Source:[3]

New results

The PICOTS-ComTeC framework provides a structured guide for formulating precise DHT definitions and selecting comparator technologies. It can be applied across contexts including study reporting, clinical and financing decisions, evidence syntheses, and regulatory or reimbursement submissions. Main domains should

always be specified, while subcategories allow optional detail. [3].

Publication 4: Mapping Digital Health Frameworks to PICOTS-ComTeC

Rationale

Health organizations at national, regional, and international levels are facilitating the innovation and integration of DHTs into healthcare systems. Numerous guidelines address evidence generation, regulatory authorization, and HTA, yet no standardised approach exists for defining and reporting DHTs (Rouleau et al. 2024; Fatehi, Samadbeik, and Kazemi 2020).

Therefore, the DH-SIG aimed to map the DHT definition items of established DHT frameworks to PICOTS-ComTeC to assess the degree of overlap, and the added value of PICOTS-ComTeC in creating a unified approach to define DHTs.

Methods

An interdisciplinary DH-SIG expert group selected 16 established national and international frameworks through online consensus discussions. These included national

HTA frameworks from Australia, Belgium, Finland, France, Germany, and the UK, as well as frameworks for standardising DHT labelling, evidence generation, reporting, or functional categorisation. One framework addressed reporting of health economic evaluations. Reviewers were trained, piloted the extraction process, and worked independently in pairs to extract general information, categorise framework purpose, and map definition items against PICOTS-ComTeC. After pairwise consensus, a third reviewer consolidated the data. Reviewers distinguished items describing DHT definitions from those describing methodological approaches. If at least one subcategory within a domain was covered, the framework was considered to overlap with that PICOTS-ComTeC domain [4].

Results

The 16 frameworks overlapped with 81% of the nine PICOTS-ComTeC domains (116/144). On average, a framework matched 7.3 domains. Across frameworks, only 48% (247/512) of PICOTS-ComTeC subcategories were covered. PICOTS-ComTeC items were unevenly

represented, with some frameworks offering greater granularity.

Table 2. PICOTS-ComTec versus DHT frameworks

	# Domains	Number of Subcategories by Domain									
		P	I	C	O	T	S	Com	Te	C	Sub-category Total
PICOTS-ComTeC	9	3	3	3	6	2	3	4	5	3	32
Comparators											
WHO CDISAH	9	3	2	1	6	1	3	4	5	1	26
CEN-ISO/TS 82304-2	7	2	1		2		1	2	3	2	13
MARS	4	1	2					3	3		9
TECH	4	2	0					1	2		5
Australia DHA	7	3	2		2		1	1	2	1	12
Belgium RIZIV	9	2	2	3	3	1	1	2	5	1	20
Finland Digi-HTA	8	2	1	2	5		1	2	4	3	20
France HAS	7	2	2		3	2	1	3	5		18
Germany DiGA	9	2	3	2	6	1	1	3	4	3	25
UK NICE	9	3	1	1	5	1	1	2	2	2	18
CONSORT-EHEALTH	9	3	2	0	2	1	1	4	3	1	17
Evidence DEFINED	6	2		1	3			3	2	1	12
iCHECK - DH	8	2	2	1	3		1	2	2	2	15
ISPOR CHEERS	7	2	0	1	2	1	2			2	10
VF-DHT	5	1			3			3	4	1	12
WHO mERA	8	2	1		1	1	1	2	5	2	15
Total	116	34	21	12	46	9	15	37	51	22	247
Mean	7.3	2.1	1.5	1.3	3.3	1.1	1.3	2.5	3.4	1.7	15.4
Max. Possible	144	48	48	48	96	32	48	64	80	48	512
% of Max. Possible	81%	71%	44%	25%	48%	28%	31%	58%	64%	46%	48%

Subcategory values are zero if the framework includes the domain without subcategories. A match on any subcategory counts as the domain being present, even if not formally defined. Cells are blank if neither domain nor subcategory is matched.

Source: [4]

New Results

Given its high overlap with established frameworks and broad coverage of relevant items, PICOTS-ComTeC can

serve as a common reference for defining DHTs across research, reporting, and HTA contexts [4].

How COVID tracing apps balance between data privacy and public health interests?

Publication 5: Recommendations regarding privacy protection and public health impact of contact tracing apps

Rationale

During the COVID-19 pandemic, most countries introduced contact tracing apps (CTA) to identify close contacts of infected individuals and prompt users to test and self-quarantine. While CTAs promised public health benefits, they also raised concerns about large-scale digital surveillance and civil rights violations (Amnesty International 2020). In response, the European Data Protection Board (EDPB) issued guidelines on anonymised mobility data for CTA developers (European Union 2020). We aimed to evaluate how existing CTAs balance between public health interests and compliance with data protection standards.

Methods

We conducted a systematic review to identify CTAs. PubMed, IEEE Access, and the ACM Digital Library were searched between January 1 and August 31, 2020, for primary research papers on CTAs. Public health potential was assessed using a checklist based on the Ada Lovelace Institute report (Ada Lovelace Institute 2020). Compliance with data privacy standards was evaluated using a checklist compiled from the European Commission's Privacy Code of Conduct for Mobile Health Apps (European Commission 2020), and EDPB guidelines (European Union 2020). CTAs were further characterised using developer and governmental websites and GitHub. Compliance with each criterion was scored +1, non-compliance -1, and missing information 0. Apps were ranked by overall score. Further methodological details are provided in [5].

Results

We identified 21 CTAs. The COVIDsafe and SwissCovid apps scored highest (15), while Alipay Health scored

lowest (-3) in achieving balance between data privacy and public health interests.

Table 2. Assessing the data privacy compliance and public health potential of contact tracing apps

	COVIDsafe	SwissCovid	Immuni	Corona-Warn-App	ABTrace Together	COCOA	Covid Watch	Smittestop	TraceTogether	ProteGo Safe	Stopp Corona	COVID Trace	Hamagen	NHS Covid-19 App	Tabaud	COVID Safe Paths	TousAntiCovid	GH Covid-19	BeAware Bahrain	Aarogya Setu	Alipay Health Code	SUM:
DATA PRIVACY																						
Voluntary basis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	15
Consent of the user is required	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1	-1	17
No mandatory measures imposed on user	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	17
Period of data retention is minimised	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	-1	0	1	1	0	16
Anonymisation	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	19
Information about gathering children's data	1	1	1	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	7
Information about data breach (prevention/planned activities)	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3
Minimisation of collected data	1	1	1	1	1	1	1	1	1	1	1	1	-1	0	-1	-1	-1	-1	-1	-1	-1	4
Centralised approach only if data minimisation ensured	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	-1	0	0	-1	0	13
Only Bluetooth with pseudo-random identifiers	1	1	1	1	1	1	1	1	1	1	1	1	-1	0	1	-1	1	-1	-1	-1	0	9
PUBLIC HEALTH INTERESTS																						
Governmental accountability	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	19
Information about efficiency threshold	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Definition of close contact	1	1	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	10
Open source code	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	14
Confirmation of positive results of COVID-19 by healthcare authorities only in a secure way (i.e. special one-time code)	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	13
Health data shared to health authorities only if anonymised or aggregated	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	18
TOTAL SCORE:	15	15	14	14	13	13	13	13	12	12	12	11	10	9	8	5	5	2	1	1	-3	

Source: [5]

New Results

Neither contact tracing app has achieved maximum score on both data privacy and public health potential. COVIDSafe and SwissCovid provided the best balance between public health interests and compliance with data protection standards.

Beyond the traditional HTA modules, what evidence is required by payers for funding decisions of DHTs?

Publication 6: Paying for Digital Health Interventions – What Evidence is Needed?

Rationale

To inform public funding decisions, traditional health technologies (e.g., pharmaceuticals) undergo HTA assessment. “HTA is a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision making in order to promote an equitable, efficient, and high-quality health system.” (O'Rourke, Oortwijn, and Schuller 2020) The European Network for HTA’s (EUnetHTA) Core Model comprises nine modules: current use, safety, clinical effectiveness,

patient and social aspects, economic, legal, ethical, organization and technical aspects (EUnetHTA 2021). For sustainable business models, DHT developers increasingly seek public funding. We aimed to explore what evidence payers require to inform reimbursement decisions of DHTs.

Methods

We conducted a systematic search in PubMed, HTA body websites, and grey literature using structured Google syntax. We sought payer evidence frameworks in countries offering public funding for DHTs. Data extraction followed the EUnetHTA core domains and DHT-specific domains such as usability, data security, interoperability and technical aspects / stability (Kolasa and Kozinski 2020). Further details are provided in [6].

Results

We identified six DHT payer frameworks. The Medical Services Advisory Committee (MSAC) framework in Australia (Kidholm et al. 2012), the Belgian National Health and Disability Insurance (RIZIV) framework (mHealthBelgium 2021), the National Institute for Health

and Care Excellence (NICE) Evidence Standards Framework (ESF) (The National Health Service (NHS) in England and Wales, the German Directory for Reimbursable Digital Health Applications (DiGA) (Federal Institute for Drugs and Medical Devices (BfArM; Germany) 2021), the evidence framework of the French National Authority for Health (HAS) (Haute Autorité de Santé (HAS; France) 2019), and the framework of the Finnish Coordination Centre for Health Technology Assessment (FinCCHTA) (Haverinen et al. 2019).

While the description of problem and choice of comparator, safety, clinical effectiveness, patient and social aspects and economic aspects were required by all frameworks, the evidence needs were rather heterogenous in the remaining domains. The German DiGA and Australian MSAC frameworks were the most comprehensive.

New results

Payers' evidence needs are diverse in the legal, ethical, and organisation EUNetHTA domains, and the DHT-specific usability, data security, interoperability and technical aspects / stability domains. The fragmented HTA landscape may hinder the international growth of DHTs.

Table 3. Evidence needs of payers for DHT funding

		Belgium RIZIV	Australia MSAC	England / Wales NICE	Germany DiGA	France HAS	Finland FinCCHTA
EUNetHTA Core modules	Health problem and comparator	✓	✓	✓	✓	✓	✓
	Safety	✓	✓	✓	✓	✓	✓
	Clinical effectiveness	✓	✓	✓	✓	✓	✓
	Patient and social aspects	✓	✓	✓	✓	✓	✓
	Economic aspects	✓	✓	✓	✓	✓	✓
	Legal aspects	✓	✓	✓	✓		
	Ethical aspects	✓	✓	✓	✓		
	Organisational aspects	✓	✓		✓		✓
DHT specific modules	Usability		✓		✓		✓
	Data security	✓	✓		✓		✓
	Interoperability	✓	✓		✓		✓
	Technical aspects / stability		✓		✓	✓	✓

Source:[6]

What do structured assessments reveal about the quality of clinical evidence and reporting of various DHTs?

Publication 7: Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews

Rationale

In evidence-based medicine, systematic reviews are traditionally placed at the top of the hierarchy, but modern perspectives emphasise that their value depends on both methodological quality and certainty of evidence (Murad et al. 2016). Literature on DHTs is expanding rapidly, with up to 15 000 publications annually (Péntek et al. 2024), raising questions about evidence quality. We aimed to assess the magnitude of effect in DBM-based interventions in the context of methodological quality and evidence certainty using a structured approach.

Methods

We systematically searched PubMed and Cochrane Library for systematic reviews of DBM-based interventions published in 2019–2020. Alongside extracting overall effect size, we evaluated methodological quality using A Measurement Tool to Assess Systematic Reviews 2 (AMSTAR-2) (Shea et al. 2017) and evidence certainty using the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach (Guyatt et al. 2008).

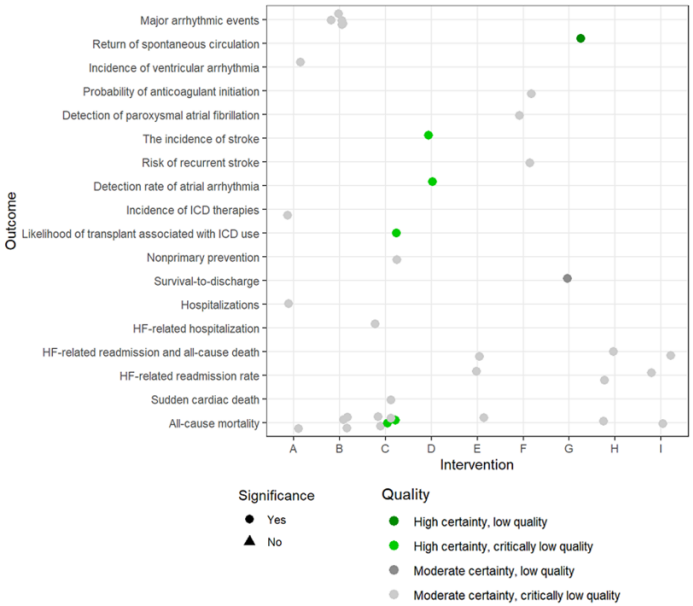
AMSTAR-2 includes 16 items and classifies reviews into four quality categories. GRADE considers five downgrading domains (risk of bias, inconsistency, imprecision, publication bias, indirectness). Based on the number of downgrades, GRADE classifies evidence certainty as high (further studies are unlikely to change our confidence in the overall effect), moderate (further research may change our confidence in the estimate or the overall effect), low (further research is very likely to change the estimate) and critically low (the estimate is very uncertain). While GRADE ratings involve expert judgement, we used uniform evaluation criteria to ensure consistency. Details are provided in [7].

Results

From 375 records, 25 systematic reviews were identified, which compared the efficacy of DBM-based and non-DBM-based interventions. Most DBMs monitored heart functions/rhythm or physical activity, for which outcomes and interventions by study were depicted in Figures 2 and 3.

Cardiac-related DBMs showed significant positive effects with high certainty for outcomes such as use of a metronome in resuscitation, detection of arrhythmias, reduced stroke risk, and lower mortality or transplant risk in ICD patients. However, methodological quality of the reviews was often low or critically low.

Figure 2. DBM interventions on cardiac function

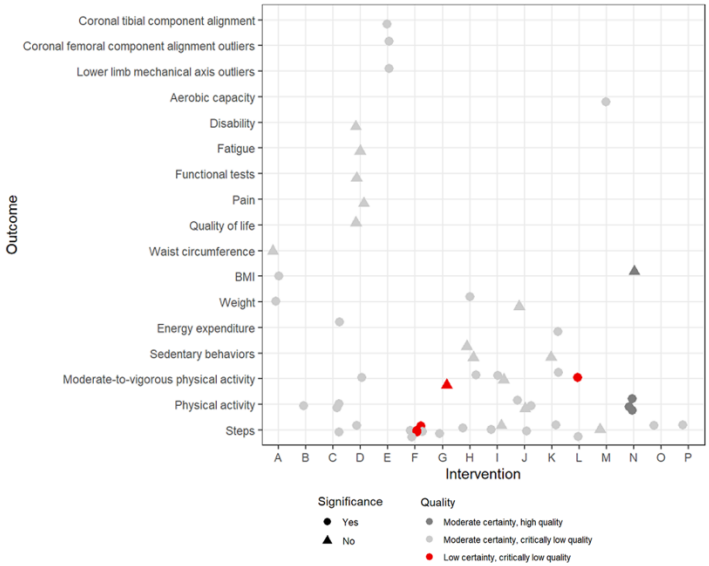


A: Cardiac resynchronization therapy, Implantable cardiac defibrillator, **B:** Fragmented QRS (fQRS) **C:** Implantable cardiac defibrillator, **D:** Implantable cardiac defibrillator, iPhone-based rhythm monitoring device, pacemakers, **E:** Impedance devices, **F:** Implantable cardiac monitor, Holter-Electrocardiogram, **G:** Metronome with a siren, **H:** Pressure sensors, **I:** Pressure sensors and Impedance devices

Source:[7]

Physical activity DBMs produced mixed results, with several negative outcomes. Yet, with moderate certainty in high-quality reviews, pedometers improved activity though not BMI. The methodological quality of reviews was low or critically low.

Figure 3. DBM interventions for physical activity



A: wristbands, smartwatches, **B:** Accelerometer, Pedometer, Yamax Digi-walker CW700, ActivPal, ActiGraph, Personal Activity Monitor, **C:** Accelerometer, pedometer, **D:** Accelerometer, pedometers, Yamax, Fitbit, **E:** Accelerometer-based navigation system, **F:** wearable activity trackers (pedometer), **G:** Activity monitor, portable tablet computers with touch screens, Fitbit, Jawbone UP24 wearable device, pedometer, accelerometer, **H:** Fitbit, **I:** Fitbit, Jawbone UP, Polar Active, Misfit Flash, Gruve Solution, LUMOback, BodyMedia Fit, SenseWear, ActiveLink, InBodyBand, **J:** Fitbit, Jawbone Up24, Combined heart rate monitor and accelerometer (Actiheart), Wrist-worn accelerometer, FIT Core, Body Media, Fitbug Orb, Polar FA20 accelerometer, **K:** Fitbit, Jawbone UP24, Gruve, LumoBack, Polar Active, Fitbug, Pebble+, Fitmeter, Personal Activity Monitor, Withings Pulse, **L:** Fitbit, Yorbody, AiperMotion, **M:** Garmin, Pedometer, Fitbit, Accelerometer, Yamax Digiwalker, Gex sensor of vital signs and smartphone, **N:** Pedometer, **O:** pedometer-based physical activity promotion, **P:** Pedometer physical activity promotion + pulmonary rehabilitation promotion

Source:[7]

New results

Cardiac DBMs show positive effects on all-cause mortality, return of circulation, stroke incidence, and arrhythmia detection with high certainty of evidence, yet the methodological quality of reviews is mostly low or critically low. In contrast, high-quality systematic reviews provide only moderate-certainty evidence that activity trackers increase step count, while evidence for broader health outcomes such as weight loss, lower body mass index, or improved quality of life is mixed or absent. Regardless of the expectations of investors and healthcare managers, or the concerns of developers, health benefits of DHTs must be demonstrated to qualify for public funding.

Publication 8: Reporting quality of machine-learning studies involving clinical populations with paediatric diabetes

Rationale

Studies using AI and machine learning (ML) in medicine are rapidly increasing, but concerns about their quality persist (Andaur Navarro et al. 2022; Coiera et al. 2018). Medical AI differs from simply running algorithms on

data: it aims to support clinical decisions and improve patient outcomes (Cabitza and Campagner 2021; Futoma et al. 2020). As such, results should be meaningful for both clinicians and computer scientists. Good reporting quality facilitates effective communication of research results between stakeholder groups, and hence, it contributes to the stabilisation of the meaning of an innovation towards its closure. Therefore, we aimed to systematically evaluate the reporting quality of primary ML/AI studies in paediatric diabetes, a clinical area rapidly adopting advanced technologies (Danne and Limbert 2020).

Methods

We conducted a systematic review of primary studies using ML/AI in a paediatric diabetes population, published in 2016-2020. An extended list general and specific AI / ML search terms was applied. Reporting quality was assessed via the Minimum Information About Clinical Artificial Intelligence Modelling (MI-CLAIM) checklist (Norgeot et al. 2020). MI-CLAIM features 17 binary and 2 polytomous questions in five domains

following the ML workflow: study design, data and optimisation, model performance, model examination and reproducibility. Each study was evaluated by a pair comprising a medical and a computer science expert, pre-trained through pilot sessions. Items were coded “yes” (sufficient detail), “unsure” (insufficient detail), or “no” (missing). Results were depicted graphically. Further details are in [8].

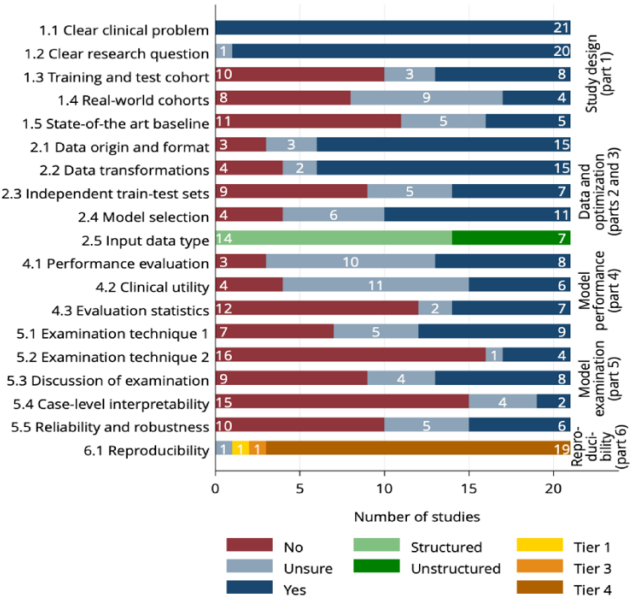
Results

From 1043 records screened, we identified 28 eligible studies, out of which 21 could be evaluated by MI-CLAIM. The studies were methodologically diverse. We found 61 unique ML techniques out of which 48 were mentioned in only one paper. The clinical problem and research questions were clearly stated in most studies, the adequate description of the training and test cohorts, performance evaluation statistics and model examination were the least often reported items. The reporting of evaluation metrics of mathematical and clinical model performance was especially vague.

New results

While problems and research questions are clearly stated, the reporting quality of AI / ML studies in paediatric diabetes must improve so that results are unequivocal and their value to clinicians is communicated in a straightforward way. Key areas for improvement are the clinical characteristics of the cohorts, clinically meaningful performance metrics and model explainability.

Figure 4. Reporting quality by MI-CLAIM items



How to select a reporting guideline for a medical AI study?

Publications 9 & 10: A systematic review of reporting guidelines for medical AI studies

Rationale

The review of reporting quality in paediatric diabetes showed that, given diverse clinical designs and study settings, MI-CLAIM was not uniformly applicable to medical AI studies [8]. Therefore, we aimed to systematically review existing medical AI reporting guidelines and explore aspects guiding their selection. A rapid review [9] was followed by an in depth analysis [10], available as a preprint, and accepted for publication in Acta Polytechnica Hungarica.

Methods

We performed a rapid review by searching PubMed in May 2022. A detailed review extended to Web of Science (WoS) and Scopus in February 2023. We extracted target audience, clinical area, study design in focus, development process according to the EQUATOR Network Guidance for Developers of Health Research Reporting Guidelines (Moher et al. 2010), type of

guideline (i.e., checklist or narrative), and whether reporting items followed IMRAD (introduction, methods, results and discussion) or the machine learning workflow used in MI-CLAIM (Norgeot et al. 2020). To assess the impact of the included reporting guidelines, we recorded Google Scholar citations. We also counted the total number of reporting items. Details are provided in [9,10]

Results

The rapid review yielded 21 studies from 424 records, while we found 24 eligible guidelines from 821 records in the detailed review. The number of citations of the guidelines were minimal compared to the number of published AI studies. Altogether, 704 reporting items were extracted, showing no consensus on the structure or content of medical AI reporting. The most robust development process was for CONSORT-AI, though it applies only to randomised clinical trials.

Table 4. Summary of medical AI reporting checklists

Characteristic	Category		T.R.U.E. (Buvat and Orlhac 2021)	(Stevens et al. 2020)	(Faes et al. 2020)	(Bates et al. 2020)	ROBUST-ML (Al-Zaiti et al. 2022)	(Cabitza and Campagner 2021)	CAIR (Olczak et al. 2021)	(Scott, Carter, and Coiera 2021)	MINIMAR (Hernandez-Boussard et al.	MI-CLAIM (Norgeot et al. 2020)	(Luo et al. 2016)	CONSORT-AI (Liu et al. 2020)	SPIRIT-AI (Cruz Rivera et al. 2020)	DECIDE-AI (Group 2021)	CHARMS (Moons et al. 2014)	CLEAR Derm (Daneshjou et al. 2022)	R-AI-DIOLOGY (Haller et al. 2022)	STREAM-URO (Kwong et al. 2021)	CLAIM (Mongan, Moy, and Kahn 2020)	Canada protocol (Morch, Gupta, and	(Schwendicke et al. 2021)	PRIME (Sengupta et al. 2020)	CLAMP (El Naqa et al. 2021)	MAIC-10 (Cerdá-Alberich et al. 2023)
Guideline development process	Development methods reported							✓		✓			✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓
	Registered in EQUATOR website			✓								✓	✓	✓	✓	✓				✓	✓	✓	✓		
	Literature review							✓		✓			✓	✓	✓	✓	✓	✓		✓		✓	✓		✓
	Delphi survey												✓	✓	✓	✓						✓	✓		
	Expert consensus meeting													✓	✓	✓	✓	✓		✓					✓
	Pilot testing													✓	✓	✓	✓						✓		✓
	Funded						✓		✓				✓	✓	✓	✓	✓			✓	✓			✓	✓
	Update policy was stated													✓	✓									✓	
	Journal / Society endorsement		✓					✓						✓							✓		✓	✓	
Target audience	Authors, reviewers, editors		✓	✓	✓			✓			✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
	Clinicians and model users				✓		✓		✓	✓	✓	✓							✓						
	Application developers																	✓		✓		✓			
Type	Narrative		✓	✓	✓	✓																			
	Checklist						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Focus	General	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
	Study design	Randomised clinical trial												✓											
		Clinical trial protocol													✓										
		Early stage clinical evaluation														✓									
		Systematic review															✓								
	Clinical area	Clinical imaging																	✓		✓			✓	✓
		Dentistry																					✓		
		Cardiovascular medicine		✓			✓																	✓	
		Cardiovascular imaging																						✓	
		Medical physics																							✓
		Mental health																				✓			
		Dermatology																✓							
		Urology																		✓					
		Neuroradiology																	✓						
		Nuclear medicine	✓																						
		Ophthalmology			✓																				
		Orthopaedics							✓																
Structure	IMRAD							✓					✓	✓	✓	✓				✓	✓			✓	✓
	Machine Learning Pipeline			✓	✓	✓	✓	✓		✓	✓	✓					✓	✓	✓				✓		
	Other		✓																			✓	✓		

New results

The diversity of reporting items suggests that medical AI/ML is still in the phase of interpretive flexibility. Reporting should reflect the AI/ML process and the primary study design. We recommend that researchers should combine a well-developed general AI checklist with established design-specific checklists for optimal reporting.

Summary of the new results

In the lack of standardized DHT taxonomies, how can DHTs be uniquely defined?

[1] Most PICO questions of DBM systematic reviews could be coded using WHO tools, but interventions on healthy populations were not classifiable. As prevention and risk reduction in healthy populations are key applications of DHTs, this gap in international medical taxonomies should be addressed to capture all use cases of DHTs.

[2] Despite the proliferation of unique digital health-related definitions over time, the concepts were

overlapping, and ambiguous about the functional domains of digital health. The DH-SIG recommended combining PICO (Population, Intervention, Comparator and Outcome) elements with specific secondary digital health terms (e.g., telesurgery, telerehabilitation). Relying on umbrella terms alone risks overly heterogenous evidence summaries of the studied interventions.
[3] The PICOTS-ComTeC framework provides a structured guide for formulating precise DHT definitions and selecting comparator technologies. It can be applied across contexts including study reporting, clinical and financing decisions, evidence syntheses, and regulatory or reimbursement submissions. Main domains should always be specified, while subcategories allow optional detail.

[4] Given its high overlap with established frameworks and broad coverage of relevant items, PICOTS-ComTeC can serve as a common reference for defining DHTs across research, reporting, and HTA contexts.

How COVID tracing apps balance between data privacy and public health interests?

[5] Neither contact tracing app has achieved maximum score on both data privacy and public health potential. COVIDSafe and SwissCovid provided the best balance between public health interests and compliance with data protection standards.

Beyond the traditional HTA modules, what evidence is required by payers for funding decisions of DHTs?

[6] Payers' evidence needs are diverse in the legal, ethical, and organisation EUNetHTA domains, and the DHT-specific usability, data security, interoperability and

technical aspects / stability domains. The fragmented HTA landscape may hinder the international growth of DHTs.

What do structured assessments reveal about the quality of clinical evidence of various DHTs?

[7] Cardiac DBMs show positive effects on all-cause mortality, return of circulation, stroke incidence, and arrhythmia detection with high certainty of evidence, yet the methodological quality of reviews is mostly low or critically low. In contrast, high-quality systematic reviews provide only moderate-certainty evidence that activity

trackers increase step count, while evidence for broader health outcomes such as weight loss, lower body mass index, or improved quality of life is mixed or absent. Regardless of the expectations of investors and healthcare managers, or the concerns of developers, health benefits of DHTs must be demonstrated to qualify for public funding.

[8] While problems and research questions are clearly stated, the reporting quality of AI / ML studies in paediatric diabetes must improve so that results are unequivocal and their value to clinicians is communicated in a straightforward way. Key areas for improvement are the clinical characteristics of the cohorts, clinically meaningful performance metrics and model explainability.

How to select a reporting guideline for a medical AI study?

[9,10] The diversity of reporting items suggests that medical AI/ML is still in the phase of interpretive flexibility. Reporting should reflect the AI/ML process and the primary study design. We recommend that

researchers should combine a well-developed general AI checklist with established design-specific checklists for optimal reporting.

Impacts of the results

The research presented in this thesis booklet has contributed to the public discourse on the health technology assessment of digital health technologies. In doing so, it is expected to support social construction processes through which diverse meanings of digital health are negotiated, and structuration processes through which reporting standards and evidence frameworks both shape and are shaped by actors. Together, these contributions can facilitate the emergence of digital health innovations that are both socially valuable and economically viable.

As of September 2025, the ten included publications were cited 174-times, out of which 145 were independent citations in MTMT. The number of Google Scholar citations is 276.

PICOTS-ComTeC has been available on the EQUATOR network, the leading international forum for medical reporting guidelines.

https://www.equator-network.org/reporting-guidelines/the-picots-comtec-framework-for-defining-digital-health-interventions-an-ispor-special-interest-group-report/

Through the PICOTS-ComTeC publications [2,3,4], the DH-SIG gained ISPOR's approval to develop CHEERS-DHI, a digital health-specific version of the established CHEERS reporting guideline for health economic evaluations.

https://www.ispor.org/heor-resources/presentations-database/session-cti/ispor-2025/advancing-the-definition-and-reporting-of-digital-health-interventions-from-picots-comtec-to-cheers-dhi

Based on PICOTS-ComTeC, ISPOR has developed an online course titled: "Planning Digital Health Apps With Evidence in Mind". The course is intended educate globally digital health developers on structured reporting and evidence-based development.

https://learning.ispor.org/topclass/expand.do?template=New_CourseHome&id=449416&activitytype=28&offeringId=448905&learningPage=TrainingHistory

The research related to the development of PICOTS-ComTeC [2,3,4] has been presented in symposia and as

research abstracts in multiple international conferences and workshops, including ISPOR Midwest Chapter (February 2021), Virtual ISPOR 2021, HTAi 2021 Virtual Annual Meeting, Virtual ISPOR Europe 2021, ISPOR Europe 2023 (Copenhagen), ISPOR 2024 (Atlanta), ISPOR Europe 2024 (Barcelona), ISPOR 2025 (Montreal), and the HPI-ZEW-DIW Workshop on Applied Economics in Digital Health (5-6 June, 2025, Mannheim)

https://www.ispor.org/member-groups/special-interest-groups/digital-health

The work on AI reporting guidelines [9,10], was presented on a Workshop on the Assessment of AI Health Technologies at the 2024 HTAi Annual Meeting (Seville)

The experiences with DHT evidence syntheses were summarised in: Péntek M, et al. 10 Pragmatic Points to Consider When Performing a Systematic Literature Review of Clinical Evidence on Digital Medical Devices. Acta Polytechnica Hungarica, 2023, 20(8):281-303.

References

Ada Lovelace Institute. 2020. 'Exit through the App Store? Rapid evidence review. London, UK: Ada Lovelace Institute', Accessed December 21, 2024. https://www.adalovelaceinstitute.org/case-study/exit-through-the-app-store/.

AHRQ. na. "Six Domains of Healthcare Quality." In. online: Agency for Healthcare Research and Quality.

Al-Zaiti, S. S., A. A. Alghwiri, X. Hu, G. Clermont, A. Peace, P. Macfarlane, and R. Bond. 2022. 'A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML)', *Eur Heart J Digit Health*, 3: 125-40.

Amnesty International. 2020. 'Digital surveillance to fight COVID-19 can only be justified if it respects human rights. Amnesty International', Accessed December 21, 2024. https://www.amnesty.org/en/latest/news/2020/04/covid19-digital-surveillance-ngo/

Andaur Navarro, C. L., J. A. A. Damen, T. Takada, S. W. J. Nijman, P. Dhiman, J. Ma, G. S. Collins, R. Bajpai, R. D. Riley, K. G. M. Moons, and L. Hooft. 2022. 'Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review', *BMC Med Res Methodol*, 22: 12.

Babrak, L. M., J. Menetski, M. Rebhan, G. Nisato, M. Zinggeler, N. Brasier, K. Baerenfaller, T. Brenzikofer, L. Baltzer, C. Vogler, L. Gschwind, C. Schneider, F. Streiff, P. M. A. Groenen, and E. Miho. 2019. 'Traditional and Digital Biomarkers: Two Worlds Apart?', *Digit Biomark*, 3: 92-102.

Bates, D. W., A. Auerbach, P. Schulam, A. Wright, and S. Saria. 2020. 'Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence', *Ann Intern Med*, 172: S137-S44.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor J. Pinch. 1993. *The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology.* (MIT Press: Cambridge, MA).

Buvat, I., and F. Orlhac. 2021. 'The T.R.U.E. Checklist for Identifying Impactful Artificial Intelligence-Based Findings in Nuclear Medicine: Is It True? Is It Reproducible? Is It Useful? Is It Explainable?', *J Nucl Med*, 62: 752-54.

Cabitza, F., and A. Campagner. 2021. 'The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies', *Int J Med Inform*, 153: 104510.

Canada Health Infoway. 2021. 'What Is Digital Health', Canada Health Infoway, Accessed 2021.04.13.

Cerda-Alberich, L., J. Solana, P. Mallol, G. Ribas, M. Garcia-Junco, A. Alberich-Bayarri, and L. Marti-Bonmati. 2023. 'MAIC-10 brief quality checklist for publications using artificial intelligence and medical images', *Insights Imaging*, 14: 11.

Coiera, E., E. Ammenwerth, A. Georgiou, and F. Magrabi. 2018. 'Does health informatics have a replication crisis?', *J Am Med Inform Assoc*, 25: 963-68.

Cruz Rivera, S., X. Liu, A. W. Chan, A. K. Denniston, M. J. Calvert, A. I. Spirit, Consort-Ai Working Group, A. I. Spirit, Consort-Ai Steering Group, A. I. Spirit, and Consort-Ai Consensus Group. 2020. 'Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension', *Nat Med*, 26: 1351-63.

Daneshjou, R., C. Barata, B. Betz-Stablein, M. E. Celebi, N. Codella, M. Combalia, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, K. Liopyris, J. Malvehy, H. S. Seog, H. P. Soyer, E. R. Tkaczyk, P. Tschandl, and V. Rotemberg. 2022. 'Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology: CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group', *JAMA Dermatol*, 158: 90-96.

Danne, Thomas, and Catarina Limbert. 2020. 'COVID-19, type 1 diabetes, and technology: why paediatric patients are leading the way', *The Lancet Diabetes & Endocrinology*, 8: 465-67.

Dosi, Giovanni. 1982. 'Technological paradigms and technological trajectories', *Research Policy*, 11: 147-62.

El Naqa, I., J. M. Boone, S. H. Benedict, M. M. Goodsitt, H. P. Chan, K. Drukker, L. Hadjiiski, D. Ruan, and B. Sahiner. 2021. 'AI in medical physics: guidelines for publication', *Med Phys*, 48: 4711-14.

EUnetHTA. 2021. 'HTA Core Model®', Accessed December 21, 2024. https://eunethta.eu/hta-core-model/.

European Commission. 2020. 'Privacy code of conduct on mobile health apps', Accessed December 21, 2024. https://ec.europa.eu/digital-single-market/en/privacy-code-conduct-mobile-health-apps.
European Union. 2020. 'Commission recommendation (EU) 2020/518 of 8 April 2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymized mobility data', Accessed December 21, 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32020H0518.
Faes, L., X. Liu, S. K. Wagner, D. J. Fu, K. Balaskas, D. A. Sim, L. M. Bachmann, P. A. Keane, and A. K. Denniston. 2020. 'A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies', *Transl Vis Sci Technol*, 9: 7.
Fatehi, F., M. Samadbeik, and A. Kazemi. 2020. 'What is Digital Health? Review of Definitions', *Stud Health Technol Inform*, 275: 67-71.
FDA. 2020. 'What is Digital Health?', FDA, Accessed 2021.04.13. https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health.
Federal Institute for Drugs and Medical Devices (BfArM; Germany). 2021. "The Fast-Track Process for Digital Health Applications (DiGA)
according to Section 139e SGB V - A Guide for Manufacturers, Service Providers and Users." In. Berlin: Federal Institute for Drugs and Medical Devices.
Freeman, Christopher. 1987. *Technology, Policy, and Economic Performance: Lessons from Japan* (Pinter Publishers: London).
Futoma, J., M. Simons, T. Panch, F. Doshi-Velez, and L. A. Celi. 2020. 'The myth of generalisability in clinical research and machine learning in health care', *Lancet Digit Health*, 2: e489-e92.
Giddens, Anthony. 1984. *The Constitution of Society. Outline of the Theory of Structuration* (University of California Press: Berkeley and Los Angeles).
Golinelli, D., E. Boetto, G. Carullo, A. G. Nuzzolese, M. P. Landini, and M. P. Fantini. 2020. 'Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature', *J Med Internet Res*, 22: e22280.
Group, Decide-Ai Steering. 2021. 'DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence', *Nat Med*, 27: 186-87.
Guyatt, G. H., A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, H. J. Schunemann, and Grade Working Group. 2008. 'GRADE: an emerging consensus on rating quality of evidence and strength of recommendations', *BMJ*, 336: 924-6.
Haller, S., S. Van Cauter, C. Federau, D. M. Hedderich, and M. Edjlali. 2022. 'The R-AI-DIOLOGY checklist: a practical checklist for evaluation of artificial intelligence tools in clinical neuroradiology', *Neuroradiology*, 64: 851-64.
Haute Autorité de Santé (HAS; France). 2019. "Medical device evaluation by the CNEDiMTS (Medical Device and Health Technology Evaluation Committee) - Guide to the specific features of clinical evaluation of a connected medical device (CMD) in view of its application for reimbursement." In. Saint-Denis-La Plaine.
Haverinen, Jari, Niina Keränen, Petra Falkenbach, Anna Maijala, Timo Kolehmainen, and Jarmo Reponen. 2019. 'Digi-HTA: Health technology assessment framework for digital healthcare services', *Finnish Journal of eHealth and eWelfare*, 11: 326–41.
Heinemann, L. 2021. 'Expenditure for the Development of a Medical Device: Much Higher Than Commonly Assumed', *J Diabetes Sci Technol*, 15: 3-5.
Hernandez-Boussard, T., S. Bozkurt, J. P. A. Ioannidis, and N. H. Shah. 2020. 'MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care', *J Am Med Inform Assoc*, 27: 2011-15.
Hu, Krystal 2023. 'ChatGPT sets record for fastest-growing user base - analyst note', Reuters, Accessed December 20, 2024. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.
ICLG Group. 2024. 'Digital Health Laws and Regulations Recent Updates on Emerging Trends in the Global Regulation of Digital Health: Fragmented Frameworks Continue Striving to Catch Up With Technological Advancement 2024', Accessed December 20, 2024. https://iclg.com/practice-areas/digital-health-laws-and-regulations/03-recent-updates-on-emerging-trends-in-the-global-regulation-of-digital-health-fragmented-frameworks-continue-striving-to-catch-up-with-technological-advancement.
IQVIA Institute for Human Data Science. 2017. "The Growing Value of Digital Health. Evidence and Impact on Human Health and the Healthcare System." In.: IQVIA.
Junger, S., S. A. Payne, J. Brine, L. Radbruch, and S. G. Brearley. 2017. 'Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review', *Palliat Med*, 31: 684-706.

Kasoju, Naresh, N. S. Remya, Renjith Sasi, S. Sujesh, Biju Soman, C. Kesavadas, C. V. Muraleedharan, P. R. Harikrishna Varma, and Sanjay Behari. 2023. 'Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology', *CSI Transactions on ICT*, 11: 11-30.

Khan, Z. A., K. Kidholm, S. A. Pedersen, S. M. Haga, F. Drozd, T. Sundrehagen, E. Olavesen, and V. Halsteinli. 2024. 'Developing a Program Costs Checklist of Digital Health Interventions: A Scoping Review and Empirical Case Study', *Pharmacoeconomics*, 42: 663-78.

Kidholm, K., A. G. Ekeland, L. K. Jensen, J. Rasmussen, C. D. Pedersen, A. Bowes, S. A. Flottorp, and M. Bech. 2012. 'A model for assessment of telemedicine applications: mast', *Int J Technol Assess Health Care*, 28: 44-51.

Kolasa, K., and G. Kozinski. 2020. 'How to Value Digital Health Interventions? A Systematic Literature Review', *Int J Environ Res Public Health*, 17.

Koonin, L. M., B. Hoots, C. A. Tsang, Z. Leroy, K. Farris, T. Jolly, P. Antall, B. McCabe, C. B. R. Zelis, I. Tong, and A. M. Harris. 2020. 'Trends in the Use of Telehealth During the Emergence of the COVID-19 Pandemic - United States, January-March 2020', *MMWR Morb Mortal Wkly Rep*, 69: 1595-99.

Kwong, J. C. C., L. C. McLoughlin, M. Haider, M. G. Goldenberg, L. Erdman, M. Rickard, A. J. Lorenzo, A. J. Hung, M. Farcas, L. Goldenberg, C. Nguan, L. H. Braga, M. Mamdani, A. Goldenberg, and G. S. Kulkarni. 2021. 'Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework', *Eur Urol Focus*, 7: 672-82.

Liu, X., S. Cruz Rivera, D. Moher, M. J. Calvert, A. K. Denniston, A. I. Spirit, and Consort-Ai Working Group. 2020. 'Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension', *Nat Med*, 26: 1364-74.

Luo, W., D. Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, A. Shilton, J. Yearwood, N. Dimitrova, T. B. Ho, S. Venkatesh, and M. Berk. 2016. 'Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View', *J Med Internet Res*, 18: e323.

mHealthBelgium. 2021. 'Validatiepiramide', Accessed December 21, 2024. https://mhealthbelgium.be/nl/validatiepiramide.

Moher, D., K. F. Schulz, I. Simera, and D. G. Altman. 2010. 'Guidance for developers of health research reporting guidelines', *PLoS Med*, 7: e1000217.

Mongan, J., L. Moy, and C. E. Kahn, Jr. 2020. 'Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers', *Radiol Artif Intell*, 2: e200029.

Moons, K. G., J. A. de Groot, W. Bouwmeester, Y. Vergouwe, S. Mallett, D. G. Altman, J. B. Reitsma, and G. S. Collins. 2014. 'Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist', *PLoS Med*, 11: e1001744.

Morch, C. M., A. Gupta, and B. L. Mishara. 2020. 'Canada protocol: An ethical checklist for the use of artificial Intelligence in suicide prevention and mental health', *Artif Intell Med*, 108: 101934.

Murad, M. H., N. Asi, M. Alsawas, and F. Alahdab. 2016. 'New evidence pyramid', *Evid Based Med*, 21: 125-7.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change* (The Belknap Press of Harvard University Press: Cambridge, MA).

Neuendorf, Kimberly A. 2017. *The Content Analysis Guidebook 2nd Edition* (Sage Publications Inc. : Los Angeles, CA).

Norgeot, B., G. Quer, B. K. Beaulieu-Jones, A. Torkamani, R. Dias, M. Gianfrancesco, R. Arnaout, I. S. Kohane, S. Saria, E. Topol, Z. Obermeyer, B. Yu, and A. J. Butte. 2020. 'Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist', *Nat Med*, 26: 1320-24.

O'Rourke, B., W. Oortwijn, and T. Schuller. 2020. 'Announcing the New Definition of Health Technology Assessment', *Value Health*, 23: 824-25.

Olczak, J., J. Pavlopoulos, J. Prijs, F. F. A. Ijpma, J. N. Doornberg, C. Lundstrom, J. Hedlund, and M. Gordon. 2021. 'Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal', *Acta Orthop*, 92: 513-25.

Péntek, M, **Z Zrubka**, L Gulácsi, M Weszl, JT Czere, and T Haidegger. 2024. '10 Pragmatic Points to Consider When Performing a Systematic Literature Review of Clinical Evidence on Digital Medical Devices', *Acta Polytechnica Hungarica*, 20: 281-303.

Pinch, Trevor J., and Wiebe E. Bijker. 1984. 'The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other', *Social Studies of Science*, 14: 399-441.

Rouleau, G., K. Wu, K. Ramamoorthi, C. Boxall, R. H. Liu, S. Maloney, J. Zelmer, T. Scott, D. Larsen, H. C. Wijeysundera, D. Ziegler, S. Bhatia, V. Kishimoto, C. Steele Gray, and L. Desveaux. 2024.

'Mapping Theories, Models, and Frameworks to Evaluate Digital Health Interventions: Scoping Review', *J Med Internet Res*, 26: e51098.
Schwendicke, Falk, Tarry Singh, Jae-Hong Lee, Robert Gaudin, Akhilanand Chaurasia, Thomas Wiegand, Sergio Uribe, and Joachim Krois. 2021. 'Artificial intelligence in dental research: Checklist for authors, reviewers, readers', *Journal of Dentistry*, 107.
Scott, I., S. Carter, and E. Coiera. 2021. 'Clinician checklist for assessing suitability of machine learning applications in healthcare', *BMJ Health Care Inform*, 28.
Sengupta, P. P., S. Shrestha, B. Berthon, E. Messas, E. Donal, G. H. Tison, J. K. Min, J. D'Hooge, J. U. Voigt, J. Dudley, J. W. Verjans, K. Shameer, K. Johnson, L. Lovstakken, M. Tabassian, M. Piccirilli, M. Pernot, N. Yanamala, N. Duchateau, N. Kagiyama, O. Bernard, P. Slomka, R. Deo, and R. Arnaout. 2020. 'Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist: Reviewed by the American College of Cardiology Healthcare Innovation Council', *JACC Cardiovasc Imaging*, 13: 2017-35.
Shea, B. J., B. C. Reeves, G. Wells, M. Thuku, C. Hamel, J. Moran, D. Moher, P. Tugwell, V. Welch, E. Kristjansson, and D. A. Henry. 2017. 'AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both', *BMJ*, 358: j4008.
Shepherd, Joanna 2018. 'Consolidation and Innovation in the Pharmaceutical Industry: The Role of Mergers and Acquisitions in the Current Innovation Ecosystem', *Journal of Health Care Law and Policy*, 12: 1-28.
Stevens, L. M., B. J. Mortazavi, R. C. Deo, L. Curtis, and D. P. Kao. 2020. 'Recommendations for Reporting Machine Learning Analyses in Clinical Research', *Circ Cardiovasc Qual Outcomes*, 13: e006556.
Stevenson, J. K., Z. C. Campbell, A. C. Webster, C. K. Chow, A. Tong, J. C. Craig, K. L. Campbell, and V. W. Lee. 2019. 'eHealth interventions for people with chronic kidney disease', *Cochrane Database Syst Rev*, 8: CD012379.
Tarricone, R., F. Petracca, and H. M. Weller. 2024. '"Towards harmonizing assessment and reimbursement of digital medical devices in the EU through mutual learning"', *NPJ Digit Med*, 7: 268.
The National Health Service (NHS, England). 'Digital Technology Assessment Criteria (DTAC)', Accessed Aug 22, 2022. https://www.nhsx.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/
Tricco, A. C., E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. J. Peters, T. Horsley, L. Weeks, S. Hempel, E. A. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M. G. Wilson, C. Garritty, S. Lewin, C. M. Godfrey, M. T. Macdonald, E. V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, O. Tuncalp, and S. E. Straus. 2018. 'PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation', *Ann Intern Med*, 169: 467-73.
US Food and Drug Administration. na. "Using the PICOTS Framework to Strengthen Evidence Gathered in Clinical Trials—Guidance from the AHRQ's Evidence-based Practice Centers Program." In. online: US Food and Drug Administration.
Wake, D. J., F. W. Gibb, P. Kar, B. Kennon, D. C. Klonoff, G. Rayman, M. K. Rutter, C. Sainsbury, and R. K. Semple. 2020. 'ENDOCRINOLOGY IN THE TIME OF COVID-19: Remodelling diabetes services and emerging innovation', *Eur J Endocrinol*, 183: G67-G77.
WHO. 2001. 'International classification of functioning, disability and health: ICF', World Health Organisation. https://iris.who.int/handle/10665/42407.
———. 2018. "Classification of digital health interventions v1.0." In. Geneva: World Health Organisation.
———. 2019. "WHO Guideline: recommendations on digital interventions for health system strengthening." In. Geneva: World Health Organisation.
———. 2020a. 'ICD-11 International Classification of Diseases 11th Revision The global standard for diagnostic health information', World Health Organisation, Accessed December 20, 2024. https://icd.who.int/en.
———. 2020b. 'International Classification of Health Interventions (ICHI) ', World Health Organisation.

Publications related to the theses

1. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Pentek M, Gulacsi L, **Zrubka Z.** Digital Biomarker-Based Studies: Scoping Review of Systematic Reviews. JMIR Mhealth Uhealth. Oct 24 2022;10(10):e35722. doi:10.2196/35722

2. Burrell, A., **Zrubka, Z.**, Champion, A., Zah, V., Vinuesa, L., Holtorf, A.-P., Di Bidino, R., Earla, J.R., Entwistle, J., Boltyenkov, A.T., Braileanu, G., Kolasa, K., Roydhouse, J., Asche, C., Redekop, K., Pfeiffer, C., Le, L., Janodia, M., Sharkawy, M., Şaylan, M., Lee, S.-S., Glynn, S., Ganguli, A., Badawy, S., Carvalho, L.S., Ernst, F., Seal, B., van Steen, C., Patel, N., Lee, H., Doe, A., Strouss, L., Angelillo, L., Patel, C., and Paul, S.: 'How Useful Are Digital Health Terms for Outcomes Research? An ISPOR Special Interest Group Report', Value in Health, 2022, 25, (9), pp. 1469-1479

3. **Zrubka Z**, Champion A, Holtorf AP, et al. The PICOTS-ComTeC Framework for Defining Digital Health Interventions: An ISPOR Special Interest Group Report. Value Health. Apr 2024;27(4):383-396. doi:10.1016/j.jval.2024.01.009

4. Champion, A., Burrell, A., Holtorf, A.P., Di Bidino, R., Earla, J.R., Boltyenkov, A.T., Tabata-Kelly, M., Asche, C., Seal, B., **Zrubka, Z.**, and team, P.I.-C.p.: 'Towards a Common Ground for Defining Digital Health Interventions, Mapping Digital Health Frameworks to PICOTS-ComTeC: An ISPOR Special Interest Group Report', Value Health, 2025

5.. Kolasa K, Mazzi F, Leszczuk-Czubkowska E, **Zrubka Z**, Pentek M. State of the Art in Adoption of Contact Tracing Apps and Recommendations Regarding Privacy Protection and Public Health: Systematic Review. JMIR Mhealth Uhealth. Jun 10 2021;9(6):e23250. doi:10.2196/23250

6. Zah V, Burrell A, Asche C, **Zrubka Z.** Paying for Digital Health Interventions – What Evidence is Needed? Acta Polytechnica Hungarica. 2022;19(9):179-199.

7. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, **Zrubka Z.** Digital Biomarker-Based Interventions: Systematic Review of Systematic Reviews. J Med Internet Res. Dec 21 2022;24(12):e41042. doi:10.2196/41042

8. **Zrubka Z,** Kertesz G, Gulacsi L, et al. The Reporting Quality of Machine Learning Studies on Pediatric Diabetes Mellitus: Systematic Review. J Med Internet Res. Jan 19 2024;26:e47430. doi:10.2196/47430

9. **Zrubka Z,** Gulácsi L, Péntek M. Time to start using checklists for reporting artificial intelligence in health care and biomedical research: a rapid review of available tools. 2022:000015-000020.

10. **Zrubka Z,** Kovács L, Nezhad HM, Czere J, Gulácsi L, Péntek M. Artificial Intelligence in Medicine: A Systematic Review of Guidelines on Reporting and Interpreting Studies. ResearchSquare Preprint 2023.