Project Acronym:	HosmartAI
Grant Agreement number:	101016834 (H2020-DT-2020-1 – Innovation Action)
Project Full Title:	Hospital Smart development based on AI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016834

DELIVERABLE

D1.7 – Report on KPIs definition

Dissemination level:	PU -Public
Type of deliverable:	R -Report
Contractual date of delivery:	31 January 2022
Deliverable leader:	Magdalini-Afroditi Chatzikou, Declan O' Byrne, Eugena Stamuli (Pharmecons Easy Access, PhE)
Status - version, date:	Final – v1.0, 2022-01-31
Keywords:	Key Performance Indicators, Patient Reported Outcomes Measures, Patient Reported Experiences Measures, Cost Consequence Analysis, Artificial Intelligence

Executive Summary

In times of value-based health care and scarce resources of the health care systems, economic impact assessment is of increasing importance. Leveraging technology in health care has the power to impact health care industries, hospitals and patients. Examples of new treatments, known as "interventions", include diagnostic capabilities, disease modelling and prediction, customized treatment plans, enhanced electronic health records and more.

The advantages of artificial intelligence ("AI") have been extensively discussed in the medical literature. AI can use sophisticated algorithms to efficiently 'learn' features from a large volume of health care data. Once trained, an AI algorithm can use the obtained insights to assist clinical practice, reducing the data processing burden. It can also be equipped with learning and self-correcting abilities to improve its accuracy based on feedback. An AI system can assist physicians by providing up-to-date medical information from journals, textbooks and clinical practices to inform proper patient care. In addition, an AI system can help to reduce diagnostic and therapeutic errors that are inevitable in human clinical practice. Moreover, an AI system can extract useful information from a large patient population to assist in making real-time inferences for health risk alert and health outcome prediction.

HosmartAI's objective is to promote an effective and efficient health care system transformation, using AI technological developments and robotics in hospitals and the primary care sector. HosmartAI will introduce an AI platform in order for core facilities to be linked, creating smart services for health care professionals, patients, information system managers and health organisation administration.

In order for the HosmartAI objective to be implemented and physicians, hospitals and health care managers to contribute to delivering patient care within a standard operational environment, certain steps need to be performed following the theory of performance management which allows an organization to optimize or mitigate its activities according to the contribution of those activities to the organization's purpose. At HosmartAI, specific key performance indicators (KPIs) will be used, covering the most common classes of health sector outcomes, in order to measure the project's performance.

The technologies which will be analysed based on clinical, social, operational and economic parameters cover the following spectrum:

- Diagnosis revolution
- Logistic efficiency
- Treatment improvement
- Surgical support
- Assistive care

This document is structured around the theoretical background of performance management linked with KPIs applied in the health care sector. KPIs contributing to the measurement of the outcomes from health care interventions are analysed and reported. The selected indicators to be used by HosmartAI pilots cover the whole spectrum of outcomes, which are the following:

1) **Clinical Efficacy/Effectiveness**
2) **Patient Reported Outcome Measures (PROMs)**
3) **Patient or User Reported Experiences Measures (PREMs)**
4) **Productivity**
5) **Economic**

The stakeholders which will utilize those measures are: patients, physicians, caregivers, hospital managers and policy makers. Each pilot will choose the applicable clinical outcomes and PROMs/PREMs instruments best suited to the technology in scope, following the recommendation of the health economics team.

In the current report, each KPI pillar is thoroughly analysed in order for the reader to be able to understand the selected indicators and their applicability within HosmartAI and medical pilots. This report will shed light on the identification and selection of the most important health care indicators and their link with the respective stakeholders.

Deliverable leader:	Magda Chatzikou, Declan O' Byrne, Eugena Stamuli, Pharmecons Easy Access (PhE)
Contributors:	Marteyn van Gasteren, Jose Villar (ITCL) Lacrimioara Stoicu-Tivadar, John Mantas, Patrick Weber (EFMI) Rosa Almeida (INTRAS) Izidor Mlakar (UM)
Reviewers:	EFMI, ITCL
Approved by:	Athanasios Poulakidas, Irene Diamantopoulou (INTRA)

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.1	2021-06-12	Magda Chatzikou, Declan O' Byrne	Initial Draft
0.2	2021-09-19	Eugena Stamuli	Second Draft
0.3	2021-11-15	Pilot Partners, PhE	Third Draft Incorporating Feedback from pilots
0.4	2022-01-03	Magda Chatzikou	Version for internal review
0.5	2022-01-30	Magda Chatzikou	Version addressing review comments
1.0	2022-01-31	Athanasios Poulakidas, Irene Diamantopoulou	Final version for submission

Table of Contents

Table of Figures

List of Tables

Definitions, Acronyms and Abbreviations

Acronym/ Abbreviation	Title
AF	Atrial Fibrillation
AGA	American Gastroenterological Association
AI	Artificial Intelligence
CAD	Computer Aided Diagnosis/Detection
CBA	Cost Benefit Analysis
CCA	Cost Consequence Analysis
CCTA	Coronary Computed Tomography Angiography
CDSS	Clinical Decision-Support Systems
CE	Capsule Endoscopy
CE	Cost Effective
CEA	Cost Effectiveness Analysis
CFA	Confirmatory Factor Analysis
CSII	Continuous Subcutaneous Insulin Injection
CT	Computed Tomography
CMA	Cost Minimization Analysis
CUA	Cost Utility Analysis
DoA	Description of Action
EBM	Evidence Based Medicine
ECHO	ECHOcardiography
ENDOPREM	Patient Experience of GI Endoscopy Questionnaire
EGD	EsophagoGastroDuodenoscopy
HEOR	Health Economics Outcomes Research
HUI	Health Utility Index
GDPR	General Data Protection
GI	GastroIntestinal
HRI	Human-Robot Interaction
HRPU	High-Risk Pregnancy Unit
ICER	Incremental Cost Effectiveness Ratio
IMU	Inertial Measurement Unit
IOM	Institute of Medicine
JLN	Joint Learning Network
KPI	Key Performance Indicator
MDI	Multiple Daily Injections
MIS	Measurement and Information Stream
NHS	National Health System
NICE	National Institute for Health and Care Excellence
PARIS	Patient Reported Indicator Survey

Acronym/ Abbreviation	Title
PCI	Percutaneous Coronary Intervention
PD	Parkinson's disease
PHC	Primary Health Care
PM	Performance Management
PREMs	Patient Reported Experience Measures
PROs	Patient-Reported Outcomes
PROMs	Patient Reported Outcomes Measures
PROMIS	Patient-Reported Outcomes Measurement Information System
PSNCQQ	Patient Satisfaction with Nursing Care Quality Questionnaire
PSQ	Patient Satisfaction Questionnaire
PTP	Pre-Test Probability
QALY	Quality Adjusted Life Year
SDM	Shared Decision Making
SF-6D	Short Form 6 Dimensions
SF-36D	Short Form 36 Dimension
SUS	System Usability Scale
SG	Standard Gamble
SUTAQ	Service User Technology Acceptability Questionnaire
TQM	Total Quality Management
TTO	Time Trade Off
UEQ	User Experience Questionnaire
UHC	Universal Health Coverage
UREM	Usability Reported Experience Measure
VAS	Visual Analogue Scale
WHO	World Health Organization
WHOQOL	World Health Organization Quality of Life

1 Introduction

1.1 Project Information

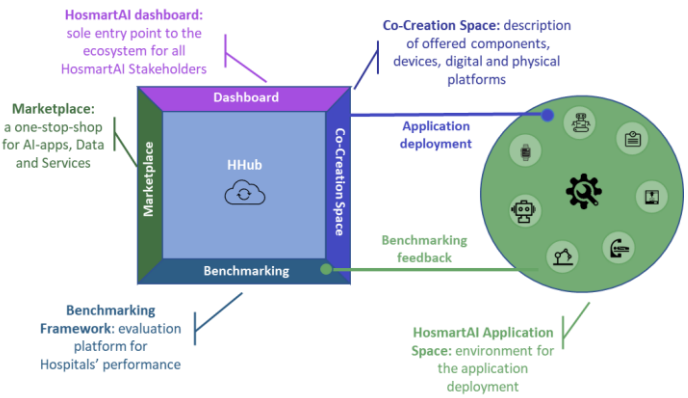
VISION

The HosmartAI vision is a strong, efficient, sustainable and resilient European **Healthcare system** benefiting from the capacities to generate impact of the technology European Stakeholders (SMEs, Research centres, Digital Hubs and Universities).

MISSION

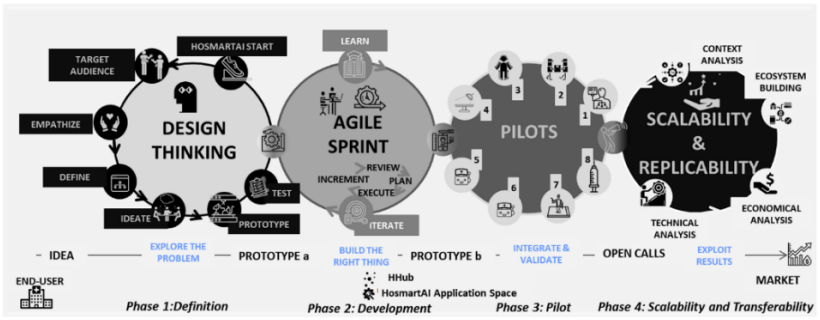
The HosmartAI mission is to guarantee the **integration** of Digital and Robot technologies in new Healthcare environments and the possibility to analyse their benefits by providing an **environment** where digital health care tool providers will be able to design and develop AI solutions as well as a space for the instantiation and deployment of AI solutions.

HosmartAI will create a common open Integration **Platform** with the necessary tools to facilitate and measure the benefits of integrating digital technologies (robotics and AI) in the healthcare system.



A central **hub** will offer multifaceted lasting functionalities (Marketplace, Co-creation space, Benchmarking) to healthcare stakeholders, combined with a collection of methods, tools and solutions to integrate and deploy AI-enabled solutions. The **Benchmarking** tool will promote the adoption in new settings, while enabling a meeting place for technology providers and end-users.

Eight Large-Scale Pilots will implement and evaluate improvements in medical diagnosis, surgical interventions, prevention and treatment of diseases, and support for rehabilitation and long-term care in several Hospital and care settings. The project will target different **medical** aspects or manifestations such as Cancer (Pilot #1, #2 and #8); Gastrointestinal (GI) disorders (Pilot #1); Cardiovascular diseases (Pilot #1, #4, #5 and #7); Thoracic Disorders (Pilot #5); Neurological diseases (Pilot #3); Elderly Care and Neuropsychological Rehabilitation (Pilot #6); Fetal Growth Restriction (FGR) and Prematurity (Pilot #1).



To ensure a user-centred approach, harmonization in the process (e.g. regarding ethical aspects, standardization, and robustness both from a technical and social and healthcare perspective), the **living lab** methodology will be employed. HosmartAI will identify the appropriate instruments (**KPI**) that measure efficiency without undermining access or quality of care. Liaison and co-operation activities with relevant stakeholders and **open calls** will enable ecosystem building and industrial clustering.

HosmartAI brings together a **consortium** of leading organizations (3 large enterprises, 8 SMEs, 5 hospitals, 4 universities, 2 research centres and 2 associations – see Table 1) along with several more committed organizations (Letters of Support provided).

Table 1: The HosmartAI consortium.

Number[1]	Name	Short name
1 (CO)	INTRASOFT INTERNATIONAL SA	**INTRA**
1.1 (TP)	INTRASOFT INTERNATIONAL SA	**INTRA-LU**
2	PHILIPS MEDICAL SYSTEMS NEDERLAND BV	**PHILIPS**
3	VIMAR SPA	**VIMAR**
4	GREEN COMMUNICATIONS SAS	**GC**
5	TELEMATIC MEDICAL APPLICATIONS EMPORIA KAI ANAPTIXI PROIONTON TILIATRIKIS MONOPROSOPIKI ETAIRIA PERIORISMENIS EYTHINIS	**TMA**
6	ECLEXYS SAGL	**EXYS**
7	F6S NETWORK IRELAND LIMITED	**F6S**
7.1 (TP)	F6S NETWORK LIMITED	**F6S-UK**
8	PHARMECONS EASY ACCESS LTD	**PhE**
9	TERAGLOBUS LATVIA SIA	**TGLV**
10	NINETY ONE GMBH	**91**
11	EIT HEALTH GERMANY GMBH	**EIT**
12	UNIVERZITETNI KLINICNI CENTER MARIBOR	**UKCM**
13	SAN CAMILLO IRCCS SRL	**IRCCS**
14	SERVICIO MADRILENO DE SALUD	**SERMAS**
14.1 (TP)	FUNDACION PARA LA INVESTIGACION BIOMEDICA DEL HOSPITAL UNIVERSITARIO LA PAZ	**FIBHULP**
15	CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE	**CHUL**
16	PANEPISTIMIAKO GENIKO NOSOKOMEIO THESSALONIKIS AXEPA	**AHEPA**
17	VRIJE UNIVERSITEIT BRUSSEL	**VUB**
18	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	**AUTH**
19	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	**ETHZ**
20	UNIVERZA V MARIBORU	**UM**

[1] CO: Coordinator. TP: linked third party.

Number[1]	Name	Short name
21	INSTITUTO TECNOLÓGICO DE CASTILLA Y LEON	**ITCL**
22	FUNDACION INTRAS	**INTRAS**
23	ASSOCIATION EUROPEAN FEDERATION FOR MEDICAL INFORMATICS	**EFMI**
24	FEDERATION EUROPEENNE DES HOPITAUX ET DES SOINS DE SANTE	**HOPE**

1.2 Document Scope

1.2.1 Introduction

Health care is one of the major success stories of our times. Medical science has improved rapidly, raising life expectancy around the world, but as longevity increases, health care systems face growing demand for their services, rising costs and a workforce that is struggling to meet the needs of its patients [REF-01] [REF-02]. Building on automation, artificial intelligence (AI) has the potential to revolutionise health care and help address some of the challenges set out above. AI can lead to better care outcomes and improve the productivity and efficiency of care delivery. It can also improve the day-to-day life of health care practitioners, letting them spend more time looking after patients and in so doing, raise productivity and improve retention. It can even get life-saving treatments to market faster. At the same time, questions have been raised about the impact AI could have on patients, practitioners and health systems, and about its potential risks; there are ethical debates around how AI and the data that underpins it should be used [REF-03][REF-04]. The current report provides the answers on the most important KPIs used in the health care sector, through the innovative AI technologies of HosmartAI and their impact on patients, physicians, academic society and health policy makers.

1.2.1.1 The Role of Key Performance Indicators (KPIs) in Health care

The quality movement concept i.e., the management of quality improvement, has spread widely during the last decades from manufacturing to service, health care, non-profit organizations, and educational institutions [REF-05]. In health care, the concept of quality has become an increasingly important factor both for patients' well-being and economic survival [REF-06] as well as among authorities, policymakers, managers, physicians, and patients [REF-07] using a number of quality management approaches, such as Evidence-Based Medicine (EBM), Total Quality Management (TQM), improvement science, professional development, and patient empowerment [REF-07]. Raven et al., (2012) posited that understanding good quality of care and how it can be measured is critical to improving health services [REF-08].

Still, defining the facets of health and quality management for health care improvement is an ongoing challenge due to individual differences and the complex relationship between health services and health outcomes [REF-09]. The Institute of Medicine (IoM), defines quality in health care as *"the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional practice."* Building on this definition, the IoM presented six characteristics of high quality care [REF-09].

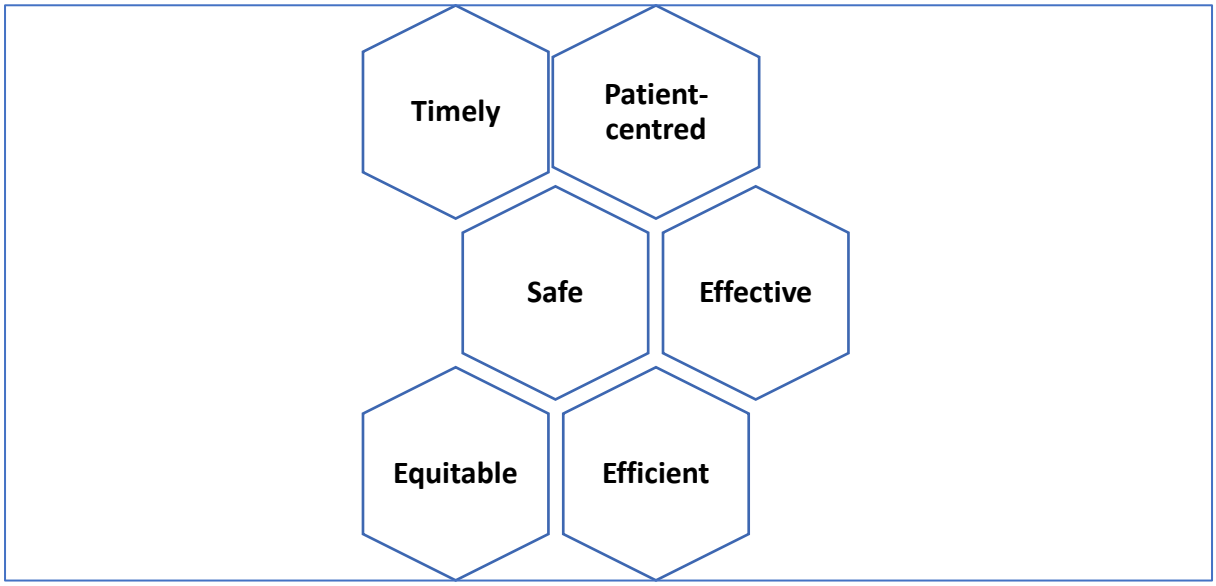


Figure 1: Characteristics of High Quality Care.

High quality care must be:

1. ***Safe***: Avoiding injuries to patients from the care that is intended to help them.
2. ***Effective:*** Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse).
3. ***Patient-centred:*** Providing care that is respectful of, and responsive to, individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions.
4. ***Timely:*** Reducing waits and sometimes harmful delays for both those who receive and those who give care.
5. ***Efficient:*** Avoiding waste, in particular waste of equipment, supplies, ideas, and energy.
6. ***Equitable:*** Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.

The goal of any health care provider is to offer an optimal level of services to their clients. Understanding the efficiency and effectiveness of organizational operations is vital to finding areas to improve the provision of services. KPIs are important in the health care environment for the interpretation of the result/objective setting, analysis of the operations and provision of evidence of effectiveness. A KPI is a measurable value that demonstrates how effectively an organisation is achieving key objectives. KPIs are classes of outcome used to measure outcome size, direction, trend, class, frequency, rate i.e., performance or preference levels for outcomes of interest to deliver objectives related to Performance Management (PM). PM is defined as an approach used by an organisation to optimize or mitigate its activities according to their contribution to the organization's purpose, the later detailed in an organisation's strategic priorities and subordinate objectives.

Through tracking health care organization's performance, the level of the provisional services can be improved and also provide a greater understanding of hospital operations. The KPI pillars in the health care setting are quadruplicate, covering mainly **a)** clinical/diagnosis/prevention effectiveness **b)** hospital/operational/cost effectiveness **c)** patients' satisfaction & quality of life improvement and **d)** productivity, which might take form such as the number of patients reviewed per annum, the cost per successfully treated patient, the cost per incremental years survival, bed occupancy rate, the rate of non-scheduled recurrent patient visits to a hospital.

Within the HOSMARTAI project, 24 partners are collaborating on health services using technological developments in robotics and AI. Using a common KPI approach will facilitate the integration of digital and robot technologies in the health care environments and enable the analysis of their benefits in all respective KPIs set by each technology.

The 24 partner organisations are divided between 8 AI and robotic pilot projects or support services. These pilot projects ("the pilots") will implement and evaluate improvements in medical diagnosis, surgical interventions, prevention and treatment of diseases, and support for rehabilitation and long-term care in several hospital and care settings. Some pilots will undertake developmental work applied to one disease, other pilots to multiple disorders. HosmartAI will target different medical aspects or manifestations such as Cancer (Pilot #1, #2 and #8); Gastrointestinal (GI) disorders (Pilot #1); Cardiovascular diseases (Pilot #1, #4, #5 and #7); Thoracic Disorders (Pilot #5); Neurological diseases (Pilot #3); Elderly Care and Neuropsychological Rehabilitation (Pilot #6); Fetal Growth Restriction (FGR) and Prematurity (Pilot #1). Figure 2 shows the role of KPIs in performance management, which is depicted as an iterative cycle of considerations. The cycle comprises statement of extant need/ opportunity, definition of a project's objectives, measures of performance and service recipients i.e., patients with the relevant condition. Furthermore, services options available to recipients, outcomes to be measured, analysis to be undertaken and review of outcomes against objectives and need/ opportunity should be defined and undertaken. Collectively this permits reflection on the value of outcomes and whether anything may additionally, likely significantly mitigate extant need or contribute to achieving extant opportunity. In such an ecosystem of inputs, outputs and outcome, a change anywhere may have ramifications through the cycle, not only on that stage's activity.

The traditional measures of productivity, i.e., efficiency, output/input and quality -adequacy of output and input- in health care and the service sector are natural considerations of performance management, but in order to be considered patient centric, patient contributions [REF-10] and involvement [REF-11] are considered necessary to establish high quality health care.

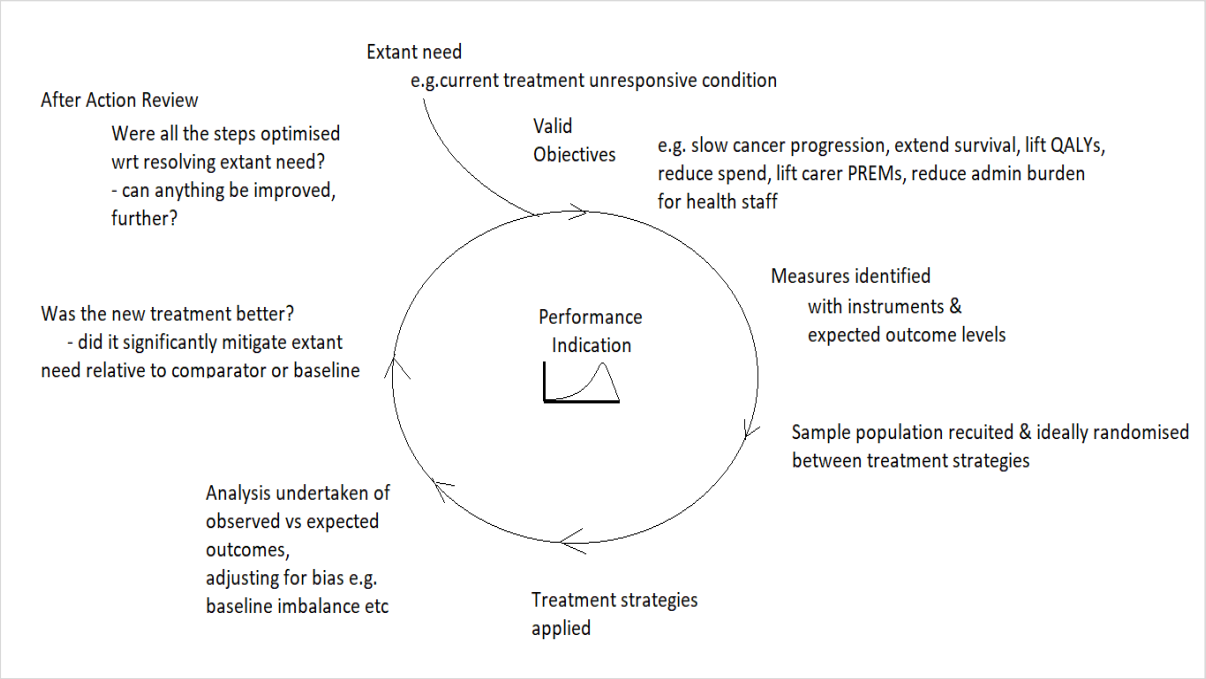


Figure 2: KPIs in an Operational/Study Ecosystem of Inputs & Outputs

1.2.2 Approach

The objective of the document is to provide a holistic analysis/view of the importance of KPIs, their application in the health care sector, mainly in the field of the use of artificial intelligence technologies. This document explores first, a description of what elements theoretically contribute to Performance Management and secondly, considerations in the application in different sectors and from large to smaller layers of organisation and technologies.

Project performance can be measured using KPIs to evaluate all projects' outcomes, determined by strategic need or opportunity and associated objectives. The value of KPIs is that in standardizing a measure, comparative studies can be made, when a comparator is used. Different combinations of KPI may be selected according to stakeholders' interests (perspective), typically decision makers. Each KPI may be a relatively direct reflection of the value of an outcome or use one of multiple choices of tool or instrument, such as questionnaires, to estimate a proxy value for that KPI. The current report covers the outcomes measurement of all involved stakeholders of the health care sector which are the following:

1. Policy Makers (Ministry of health, hospital management, EU Health care officials).
2. Physicians and health care professionals involved in medical decision-making process
3. Patients, caregivers and patient advocacy groups.
4. Academic society.

The identification of the health care related KPIs was based on a thorough literature review to be able to capture the perspective of all stakeholders of the health care sector, covering the important outcomes pillars of:

- The health care systems which are medical, productivity and economic.

- The patients and caregivers, which are patient reported outcomes and patient reported experience measures.

All AI technologies of HosmartAI have been separated in the specific KPI pillars applicable for each technology and will be analysed incrementally according to use of comparators by the projects to compare the effect difference of the new technology with respect to the comparator.

A challenge in the identification of KPIs for HosmartAI technologies was in part due to the diversity of technologies involved, where diverse outcomes required a variety of instruments for one or more KPIs for each specific technology. Additionally, multiple types of outcome, disease area, stakeholder, countries and study designs added to the complexity of finding a common theme upon which to evaluate the different projects and their technologies. Ultimately, the economic evaluation enabled a level of comparison among different endpoints, while recognising projects with different objectives, service recipients and settings are less suitable for comparison than projects which include their own internal comparators. The research team concluded that the economic analysis of choice in order to be able to match most – if not all – of the above mentioned challenges was the Cost Consequence Analysis (CCA). Cost consequence analysis enables the comparison of performance of health technologies with different endpoints by estimating the incremental cost of each AI technology. The deliverable, the D1.7 KPIs definition report, describes why performance management and KPIs are used to measure economic and social activity, including health care; who the respective stakeholders are and identify the respective facets applied in Artification Intelligence technologies based on the 8 pilots of HOSMARTAI.

1.3 Document Structure

This document is comprised of the following chapters:

Chapter 1 presents an introduction to the project and to the document, which mostly addresses the introduction of PM and KPIs applied in the health care environment. It introduces a consideration of framework for clinical and cost-effectiveness analysis and introduces the use of KPIs, in addition to describing the potential for iterative improvement of HosmartAI technologies and the stakeholders involvement in pilot studies.

Chapter 2 covers the stipulation of meaningful generalisable assessments in ethically approved studies to determine the clinical and cost-effectiveness and safety of new treatments, expressed in terms of both absolute and incremental difference. It introduces the types of health care quality measures by defining individually the structural, process and outcome measures. Additionally, the role and involvement of patients in the decision making process is analytically presented.

Chapter 3 provides a description of the KPIs used in HosmartAI, including a) Clinical effectiveness and safety b) Patient Reported Outcomes, c) Patient/User Reported Experience Measures (PREMs/UREM), d) Hospital/Health care Productivity Measures and e) Economic Outcomes. An introduction to each KPI's background, with a description of estimation and choices of instrument.

Chapter 4 presents for each pilot and for each intervention by condition a description of the pilots' objective, the range of KPIs the pilot will use and what the KPI will measure. From this, an understanding of the objectives to measure the expected type of outcome has been described.

The last, **Chapter 5**, summarises what this report has described, namely the link between performance, extant need/ opportunity and the specification of the type of measure and what that type of measure will record. It also touches upon the integration of AI and robotic technologies in the existing care framework.

2 Measuring the Effectiveness of Medical Interventions

European longevity improvements are predicted to raise the >65-year-old population from 19% in 2015 to 30% by 2050, requiring planning for the care requirements of the future [REF-12]. Health care interventions and comparators are used to determine each treatment strategy's effectiveness and cost. The extra effect per extra unit cost associated with intervention minus comparator costs and effects determines cost-effectiveness. Measures of effect are typically a clinician's most significant measures to manage a disease e.g., blood pressure to manage cardiovascular disease, but may vary depending on the nature of the intervention.

Research into interventions, permitting at least the maintenance of public health, need to be evaluated using rigorous scientific and management methods. Public health interventions should try to optimize the delivery of both efficiency and effectiveness to avoid being usurped by opportunity cost considerations i.e., the next best alternative method of using funding, such as usual care. Thus, evaluation designs must recognize and integrate the requirements of funding agents, ensure that intervention benefits can be accurately measured and conveyed, and ensure that areas for improvement are identified as needed, against extant opportunity or need. There are minimum requirements which every health care system should meet equitably: access to quality services for acute and chronic health needs; effective health promotion and disease prevention services; and appropriate response to new threats as they emerge (emerging infectious diseases, growing burden of non-communicable diseases and injuries, and the health effects of global environmental changes)[REF-13].

The assessment of health care outcomes for interventions whether a drug, procedure or technology, has evolved from observation of their value at a point in time, to assessment over a time course. The comparison with an alternative treatment strategy (the "comparator"), which is typically usual care, the prediction of future outcomes and quantification of predicted outcome uncertainty, typically reserved for studies of chronic diseases, with interventions which are not both higher in cost and lower in effect than the comparator. Where an intervention has indicated proof of principle, ideally over a comparator in pilot studies, it may warrant a Randomised Control Trial ("RCT") to determine cost and effect attributes i.e., size, direction, presence, frequency, etc, the detection of which is facilitated with a larger sample population found in RCTs, rather than in smaller studies e.g., pilot studies. The relative merit of a new treatment strategy, the intervention, compared to the comparator, is referred to as the incremental outcome or incremental difference of the intervention effect minus the comparator effect, where:

Incremental difference = mean intervention outcome (new technology) – mean comparator outcome (existing technology)

The two treatment populations, assuming a comparator is used, are compared based on the probability that the intervention is effective in either preventing, reducing the severity, or delaying an undesirable health outcome or the converse for a desirable health outcome.

In lieu of operational research, the probability of deriving the sought health outcome usually is based only on the clinical efficacy of the intervention.

Studies without a comparator may be used to demonstrate technological success rather than incremental clinical effect as without a comparator, there are limited ways to determine whether a population treated with an intervention has a higher or lower outcome. However, in some cases, historical studies of a comparator, usual care have been used as the comparator.

Estimation of outcomes (and incremental outcomes) is possible through a wide diversity of tools and approaches. We have identified five classes of outcomes relevant to health care-oriented projects seeking to determine their value to the health care system. HosmartAI study pilots were presented with a choice of up to 5 classes of KPI to be used including: clinical-effectiveness, patient reported outcome measure, patient/user reported experience measure, productivity and economic outcomes or the classes of outcome and instrument they found the best natural fit for their study. For example, local measures perhaps sponsored by a clinician may be especially clinically informative. However, their use alone i.e., without a comparator, often on small samples or in short studies can make it difficult to judge the incremental benefit and the "generalisability" of a treatment i.e., the use of the same treatment in a different setting, for the same condition to exhibits the same effect, bearing in mind risk factors and distribution of effects. In other words, an intervention without evidence to support generalisability, because of the dissimilarity of measures and outcomes can prevent an intervention's use from being supported, where direct comparisons are necessary, in which usual care would most likely be the default.

2.1 Why Measuring Health Care Outcomes is Important

Quality measurement can be a costly and time-consuming activity. Accordingly, judicious selection of indicators contributing to the aggregate understanding of health-care quality is imperative. If indicator development is to be systematic, targeting areas where the need is greatest, needs must be identified and classified by type of measure. In the HosmartAI project, the effort posed was to identify and classify clinical indicators currently being used in various countries to measure the quality of care provided by hospitals, and to identify commonalities in measurements in order to be able to apply to HosmartAI technologies. Specifically, we aimed to identify and classify indicators according to:

- **Domain** to which they apply (hospital-wide, surgical and non-surgical clinical specialties and departments).
- **Aspects of care provision** (structural, process improvement, outcome).
- **Dimensions of quality** (safety, effectiveness, efficiency, timeliness, patient-centredness and equity). This information was then used to identify gaps in current measurement [REF-14].

Health outcome estimates inform:

- funding decisions for new treatment when compared to the best alternative treatment

- safety – recording serious events e.g., death or disablement from an intervention or adverse events which may only become apparent with longer use
- public accountability by conveying treatment strategy outcomes and recording the justification for funding decisions and
- the creation of a spirit of competitiveness and application whereby a research body may create a better or alternatively applied product, service or process than other companies.

All of these applications of measurement of health outcomes seek to improve or maintain the health of patients.

Estimation of outcomes (and incremental outcomes) is possible through a wide diversity of tools and approaches. We have identified five classes of outcomes relevant to health care–oriented projects seeking to determine their value to the health care system. HosmartAI study pilots were presented with a choice of any number of 5 classes of KPI to be used from: clinical-effectiveness, patient reported outcome measure, patient/user reported experience measure, productivity and economic outcomes or the classes of outcome and instrument they found the best natural fit for their study. For example, local measures perhaps sponsored by a clinician may be especially clinically informative. However, their use alone i.e., without comparator, often on small samples or in short studies can make it difficult to judge the incremental benefit and the "generalisability" of a treatment i.e., the generalisability of a treatment is, that the same treatment in a different setting, for the same condition, exhibits the same effect, bearing in mind risk factors and distribution of effects. In other words, an intervention without evidence to support generalisability, because of the dissimilarity of measures and outcomes can prevent evidence informed policy and clinical decision making, where like with like comparisons are necessary.

2.2 Important outcomes for health care systems

Health care systems seek health improvement in outcomes at a sustainable cost or the maximum of outcome for an acceptable level of cost or extra cost. In both cases, the patient would benefit from maximisation of value, where the level of value may be defined as a ratio of additional or incremental effect to additional cost incurred from using a new treatment less that of a comparator. The reciprocal of this ratio, that is the ratio of incremental cost/incremental effect is known as the incremental cost-effectiveness ratio ("ICER"). Intervention ICERs are one of the considerations health technology agencies use in deciding on whether a new treatment represents value for money. New treatment ICERs falling below an ICER threshold, also referred to as a willingness to pay ("WTP") threshold, which is the level of extra cost per extra unit effect a funder will pay for a new intervention, are judged of sufficient value to be recommended for use, all elements being equal [REF-15].

For health care system, health care outcomes inform decision makers with an estimate of the cost-effectiveness of a new treatment permitting them to derive a judgement of worth or value from the treatment's relative performance, where a comparator or baseline is included, informing them of a relative or temporal change in value i.e., performance change from a different treatment or simply with the passage of time. Such information along with cost or

economic effect, helps a decision maker determine the worth or value of undertaking a new activity e.g., permitting a new intervention onto health care systems' funding scheme. In value-based care, which is defined by Teisberg et. Al 2020 as "the measured improvement in a person's health outcomes for the cost of achieving that improvement" is the holistic patient-centred approach that health care systems should follow [REF-16]. Understanding outcomes is central in providing value and represents an opportunity for redefining patient care. Value is created by improving the outcomes of patients with a particular clinical condition over the full cycle of care, which normally involves multiple specialties and care sites. To be successful, a key aspect of value-based care is working as teams (integrated practice units) centred around the patient's clinical condition. As medicine has become more specialized and more complex, multidisciplinary communication and trust among the care team are paramount in providing value to patients.

2.2.1 Types of Health Care Quality Measures

Measures used to assess and compare the quality of health care organizations according to Donabedian are classified as either a structure, process, or outcome measure [REF-17]. Donabedian's three components approach for evaluating the quality of care underpins measurement for improvement. Measurement for improvement has an additional component – balancing measures. Donabedian believed that structure measures have an effect on the process.

2.2.2 Structural Measures

Structural measures concern a health care provider's capacity, systems, and processes to provide care. For example:

- Whether the health care organization uses electronic medical records or medication order entry systems.
- The number or proportion of board-certified physicians.
- The ratio of providers to patients.

2.2.3 Process Measures

Process measures indicate what a provider does to maintain or improve health, either for healthy people or for those diagnosed with a health care condition. These measures typically reflect generally accepted recommendations for clinical practice. For example:

- The percentage of people receiving preventive services (such as mammograms or immunizations).
- The percentage of people with diabetes who had their blood sugar tested and controlled.

Process measures can inform consumers about the medical care they may expect to receive for a given condition or disease and they, the patient, can contribute toward improving health outcomes. Most of the health care quality measures used for public reporting are process measures.

2.2.4 Outcome Measures

Outcome measures reflect the impact of the health care service or intervention on the health status of patients. For example:

- The percentage of patients who died because of surgery (surgical mortality rates).
- The rate of surgical complications or hospital-acquired infections.

Outcome measures may seem to represent the "gold standard" in measuring quality, but an outcome is the result of numerous factors, many beyond providers' control. Risk-adjustment methods—mathematical models that correct for differing characteristics within a population, such as patient health status—can help account for these factors [REF-17].

For health care system, health care outcomes inform decision makers with an estimate of the cost-effectiveness of a new treatment permitting them to derive a judgement of worth or value from the treatment's relative performance, where a comparator or baseline is included, informing them of a relative or temporal change in value i.e., performance change from a different treatment or simply with the passage of time. Such information along with cost or economic effect as it is termed, helps a decision maker determine the worth or value of undertaking a new activity e.g., permitting a new intervention onto health care systems' funding scheme.

Many existing measures were designed to meet the needs of health care providers and health plans, which use detailed indicators to pinpoint and fix specific problems with the care they deliver. We have sought to use the most "important" metrics related to each technology, covering -as much as possible- the holistic perspective of the health care system, hospital and patient, for the HosmartAI technologies.

2.3 The Role of Patients in Decision Making Process

Identifying, clarifying and taking into account a patient's preferences for care is good clinical practice because it honours ethical principles of respect and autonomy. Also, the health benefits associated with accommodating patient reported outcomes and experiences measures will be of use in determining areas to improve, areas already sufficient and with the use of different instruments, different aspects of patient satisfaction and experience with a treatment strategy. Patient reported experience measures utility may be realized under conditions of relative equipoise [REF-18] – that is, the outcomes are sufficiently similar clinical and cost-effective enough so that one person might choose option A, whereas another would choose option B –, patient recorded outcomes e.g., experience or satisfaction measures PROs can contribute to the choice of treatment strategy, for which an organization seeks to gain regulatory approval. What counts as the 'best' outcome may be additionally take into account other measures, although patient preference, will and should remain the primary arbiter of treatment in operational practice. Shared decision-making (SDM) is promoted as an ideal model for treatment decision making [REF-19][REF-20]. Even if not well-defined, the key principle of SDM is a process that involves at least two participants (the patient and the physician), and often more (including family or professional colleagues) [REF-19]. Research

suggests that most patients want to be involved in medical decision-making and know about treatment alternatives [REF-21].

However, individual vary in the extent to which they want to participate [REF-22] [REF-23]. Therefore, patients should exercise the degree of control they wish. In some cases, patients want a large role, yet in other cases, they may delegate most decisions to a clinician. However, patients' rights to be informed and participate in decision-making is well accepted, but not always well implemented [REF-24][REF-25].

Towle and Godolphin, in 1999 [REF-26] developed a set of competencies for physicians, policy makers and patients to engage in SDM. Additional steps can involve other health professionals and form a team around the patients and including family members and others. However, this can differ depending on culture, social status, and age groups. From the study the authors acknowledge that the patient also must be competent to engage in SDM, such as defining a preferred role in decision-making, engaging in partnership with physicians, articulating health problems and expectations, communicating, accessing and evaluating information, and negotiating and agreeing on action plan [REF-27]. Therefore, SDM is a mutual process, in which both patients and professionals must be active and involved.

3 Linking Outcomes with Key Performance Indicators in Health care

3.1 Clinical Effectiveness & Safety

HosmartAI pilots' outcomes will be measured with 5 KPIs which represent the interests of different stakeholders. Clinical effectiveness could be considered the motivating force behind a study as unsurprisingly clinicians seek to administer either better or less adverse (side effects/adverse events) treatment, which are effective, efficient and safe. Patients have broadly similar hopes although likely with less interest in Pareto gains e.g., less willing to live with side effects such as a raised probability of cancer of the jaw for bone remineralisation especially where the gains may have wide margins of uncertainty or be realised only in small subpopulation. Thus, measurement of the size of clinical effect, uncertainty (standard error), its cost, time scale, and any adverse events is of interest.

The clinical outcomes may include any outcome a clinician considers of use in the management of patient's condition. These may include the accuracy of a test, survival length, mean time between recurrence or progression to a qualitatively distinct worse/better state of health e.g., disease free or diagnosed states of a disease, accuracy of diagnostic tools, minimisation of variation between repeat events, identification and assessment of presence/absence/levels of risk factors and their significance in predicting outcome, or simply saving time on the administration associated with a process e.g., reviewing diagnostic video to determine the appropriate clinical approach. The size of the effect is measured over the duration of treatment for the treatment strategy and typically compared with a control or comparator, e.g., usual care where used, such as in a randomised controlled trial. For the HosmartAI technologies, the eight pilot projects will choose, gather and analyse their own KPI measures including of clinical effectiveness.

A patient's course of health may be conceived as a continuum through life, described either by complex algorithms of varying patterns of health inputs and outcomes associated with a condition or alternatively, a patient's course of health may be considered as a transition through various discrete states of health. These states may cover the period from entry to a trial or diagnosis to the end of an observation period i.e., a trial and extrapolated to the expected cessation of change in effect e.g., at death or for example, recovery from an infection, surgery or a sprained ankle. In the discrete states approach, known as a state transition model [REF-28] e.g., for cancer, a patient may be said to transition from progression free survival to a progressed state to death. There may be variations in the route. While for infections, at their simplest level, a model may be composed of the states: susceptible, infected, recovered. Model variations could include exposed and reinfected states with various possible routes between states. Model parsimony is a useful discipline, enforcing hypothetical clarity on what elements constitute significant states, a necessity for model construction. Costs, quality of life and risk rates for clinical effects within states and rates of transition between states with some other considerations allow, through the assignment of input parameters, to state an extrapolation of the model outcomes, identified in a trial. This process permits estimation of the level of long-term inputs, outcomes and probability of cost-

effectiveness through stochastic parametric variation, and the probability of cost-effectiveness.

Clinical effectiveness may track any activity thought to lend itself to an improvement in patient outcomes, although traditionally these activities and associated outcomes have tended to comprise diagnoses, measures or estimates of mortality, prognoses and biochemical, electrophysiological or other observed markers of the process of a functioning body. Likewise, the measurement and reporting of adverse events is important so that areas needing further research are identified. The essence of the measurement is to get an estimate of how well or poorly a process works as a proxy for how a patient is faring. Typically, this would be compared to existing process or treatment e.g., Usual Care as the comparator to allow removal of background effects where an intervention was applied concurrent with a comparator or provide an alternative treatment strategy consideration where the two treatment strategies were each applied alone on a split sample population.

Aside from traditional interventions, additional services may also assist both patient and staff and these may require different outcomes to be measured. For example, compare the "treatment" strategies of receptionist Usual Care or an AI appointment scheduling algorithm, used, for illustrative purposes, around a busy hospital reception/administration desk, managing incoming and exiting day-care appointment patients and their digital notes from different departments. These two treatment strategy approaches may show differences in the speed of booking an appointment, the number of patients they can handle per day, the patient and staff experience through the intervention vs comparator and the fit to multiple and potentially departmentally varying priorities for resources and clinical areas with some parameters independent and others interacting; in short, a complex task, for which a suitably well programmed AI algorithm may be ideally suited to constant demand.

Other departments may have different outcome needs. For example, in a department using tissue structure to identify malignancies or pathology; accuracy, productivity, reproducibility and patient survival and or quality of life may show varied outcomes between intervention and comparator. For example, the comparator, Usual Care i.e., clinician undertaken or the intervention, AI algorithm, analysis of fused tissue imagery with other forms of data e.g., genotypic may help improve accuracy, the number of correctly diagnosed cases of all potential disease cases seen by that clinic and the non-diagnosis of cases who do not have the disease of all people presenting without the disease. Mapping tissue samples and subsequent AI imagery analysis may permit more efficient analysis in the future i.e., more analysis per unit time. Furthermore, minimized variation between expected and observed outcomes across serial analysis would permit improvements in accuracy, with potentially further benefits as dose refinement, fewer side effects/better safety, better treatment outcome and quality of life expectations in patients. However, the recent COVID pandemic has also highlighted the risks of a just in time marketplace with limited internal resilience. Of the two treatment strategies, the traditional approach ensures a supply of competent future surgeons, although whether their numbers alone would meet the needs of expanding aging population is another matter.

3.2 Patient Reported Outcomes Measures (PROMs)

PROMs are tools formulated as a series of health-related questions to evaluate individual patient health status such as when comparing different treatment strategies for their impact on patient health states [REF-29]. PROMs are self-reported instruments and consequently, are a direct reflection of the patient's assessment of their capacity to experience or undertake the subject of a PROM's questions, rather than an interpretation of patient reflections by a third party e.g., a clinician or carer, unless the patient is unable to respond, then the carer may answer on behalf of the patient. PROMs ask patients to answer a typically limited number of questions, where the patient selects a level of experience most closely reflecting their current health status at the time of answering. The level reflects the patient's lesser to greater difficulty conducting or experiencing the subject of the question. The summed answers, subtracted from 1, yield the health status e.g., quality of life and when treatment duration is weighted by QoL, quality of life weighted years also known quality adjusted life years ("QALYs") are estimated. QALYs are a measure combining both the level and duration of quality of life for a patient. Accordingly, PROMs allow impact assessment of various treatment strategies on patient health outcomes to be assessed by patients and over time, on patients [REF-29].

PROMs facilitate patient self-assessment of the impact of a treatment across different domains of their life from a capability perspective i.e., the degree of a patient's ability to conduct that activity.

PROMs primary use is to measure patient's quality of life in qualitative terms. They are used in cost-effectiveness evaluations to determine health economic outcomes i.e., QoL during treatment with different treatment strategies over time. When utility measures are used, the evaluation is referred to as a cost utility analysis [REF-31]. A utility measure reflects individual preference for different health outcomes e.g., life and death, measured on an interval scale with 0 reflecting death and 1 full health. During the years, QoL has become the basic metric for the reimbursement decision making process in medicine, both pharmaceuticals and medical technologies [REF-33][REF-34].

3.2.1 Patient Centred Approaches

In medicine, patient-centred approaches are used where the consequences of conditions can be better understood from an experiential, subjective experience, rather than a theoretical perspective of experience [REF-35][REF-36]. For example, for dysarthria patients, affected by weak or poorly controlled speech muscles, a speech-language pathologist may be the best assessor of impairment risk factors related to velopharyngeal dysfunction and their contribution to dysarthria. Yet the patient is best suited to assess impact on psychosocial factors such as their participation in various professional, social or personal activities [REF-29]. Therefore, both clinical-effectiveness and PRO measures are needed to identify likely treatments outcomes for both clinical and quality of life improvement.

3.2.2 PROMs in Economic Analysis

In cost-effectiveness or cost-utility analysis, incremental cost-effectiveness ratio ("ICER") is estimated using the cost difference as a numerator divided by the effect difference as denominator. The effect is typically measured in both clinical effectiveness measures and with PROMs, normally as QALYs in a CUA. However, a variation of theme exists whereby where if no primary outcome difference or QALY difference is noted, a cost minimization strategy might be undertaken to identify the treatment with the lowest cost i.e., why pay more if there are no further clinical of health economic consequences, unless there were significantly different safety or adverse events or side effects with one treatment over another. For example, a comparison of surgical techniques for myocardial revascularisation examined off- and on-pump coronary bypass surgery for myocardial revascularisation in Greek adults. 60 off-pump and 42 on-pump patients were assessed for all-cause mortality, cost-effectiveness and QoL [REF-37]. No significant difference in the primary outcome, in-hospital mortality, was found between the treatments, although the off-pump group had significant and greater reduction in complication rate - the off-pump group was 41% vs 72% for the on-pump group, p=0.001. The off-pump group had a lower cost than the on-pump group i.e., costs - off-pump €6.52 +/-0.93 vs on-pump €9.87 +/- 1.30, p=0001; the off-pump group had shorter hospital stay duration than the on-pump group - mean length of stay for off-pump group 4.93 +/-0.93 days, on-pump, 6.58 +/- 1.04 days, p0.0001. One year post surgery, quality of life was not significantly different between treatments. Such a study with equal primary clinical outcomes i.e., equal survival between treatment strategies and no significant difference in QoL would be suitable for a cost-minimisation analysis. The lower complication rate, costs and length of stay for the off-pump treatment would suggest the off-pump rather than on-pump treatment would be the better choice, although QoL during the first as well after 1 year would need to be reviewed, as well as treatment longevity.

Furthermore, the use of PROMs for individual patient management has been consistently shown to improve clinician-patient communication, detection of problems, management of conditions and patient outcomes, such as symptom control, health-related quality-of-life, and functioning [REF-38][REF-39].

3.2.3 Valuation

PROM health states are accorded values by survey of either the public for generic PROMs or patients with the relevant condition for condition specific PROMs. Different valuation methods are available including Time Trade Off (TTO), Standard Gamble (SG) and rating scales also known as Visual Analogue Scale (VAS)[REF-40][REF-41][REF-42]. Time trade-off [REF-31] is a method of eliciting health state weights by asking patients to trade Y years in health state X versus Z years in full health e.g., 20 years at 0.5 QoL in X versus 12 years full health in state Z. Z is varied until the respondent is indifferent between the alternatives e.g., 20 years at 0.5 QoL in X versus 16 years full health for Z. The quality weight of X is then set equal to Z/Y, in this example 16/20 equals 0.8 [REF-43][REF-44] [REF-45].

Standard gamble (SG) weights for health states are determined by offering respondents in health state X, a defined number of years for certain or to choose between 2 reference outcomes, typically full health for the same number of years or immediate death. The

probability of full health is varied until the respondent is indifferent between the two alternatives i.e., full health at probability 0.28 for 12 years or health state X at QoL 0.64 for 12 years. The indifference probability is the weight assigned to state X [REF-45].

For the rating scale, a respondent marks their health state to be assessed on a 0-100 "thermometer" scale, where 0 = immediate death and 100, full health. The weight assigned to the health is the reading divided by 100 [REF-45].

Discussion on the methods relative performance and benefits may be found at Torrance (1976) and Drummond (1987) [REF-31][REF-46].

3.2.4 Types of PROM

There are many sorts of PROM which may be categorized (University of Oxford PROM Group, accessed 2021) as:

- **Condition-specific e.g.,** the Rotator Cuff-Quality Of Life (RC-QOL), cancer specific EORTC-QLQ-C30
- **Population-specific e.g.,** Child Health and Illness Profile – Child Edition (CHP-CE), Child Health Utility -9 Dimension (CHU-9D)
- **Dimension-specific e.g.,** Hospital Anxiety and Depression Scale (HADS)
- **Generic/General Population e.g.,** SF-36, EQ5D, HUI
- **Individualised** e.g., Patient Generated Index
- **Summary items e.g.,** UK General Lifestyle Survey questions about accidents
- **Utility measures e.g.,** EuroQol, Health utilities index (HUI), SF6D

For example, for a condition specific PROM, the European Organisation for Research and Treatment of Cancer has created instruments e.g., EORTC-QLQ-C30 Quality of Life of Cancer Patients [REF-47]. With EORTC, patients select between not at all, a little, quite a bit, very much for the questionnaire's domains. Using cut-offs in the EORTC-QLQ-C30 outcome, another study was able to identify patients with unmet needs, helpful for clinical management [REF-38][REF-39][REF-40].

In addition to a PROM being categorized as one form or another of the preceding list, multiple PROMs may be used to determine where or whether a patient is (most) affected by some treatment strategy above another. For example, for atopic dermatitis (AD) and chronic hand eczema (CHE), the condition specific PROMs Dermatology Life Quality Index, Pruritus/Itch Numeric Rating Scale, Patient-Oriented Eczema Measure, and Quality of Life in Hand Eczema Questionnaire appear frequently in the literature [REF-48], even at the risk of some overlap [REF-41][REF-42].

Alternatively, generic PROMs could be used instead or additionally PROMs such as EQ-5D [REF-49][REF-50], WHOQOL [REF-51] and HUI [REF-34] are all utility measures which measure QoL in the general population. Utility measures measure the quality of life, permitting comparison across diseases e.g., Health Utilities Index (HUI) [REF-34]. HUI is formatted as HUI-2 with 24,000 unique health states or HUI-3 with 972, 000 unique health states. HUI measures sensation, mobility, pain, cognition, ambulation and emotion. It can be used to calculate

QALYs. Short Form 6 Dimensions (SF6D), a Short form 6 dimension of a truncated version of SF36, is suitable for deriving QoL [REF-52][REF-53]. Rather than SF36, which is unsuitable alone for creating a value set because it does not allow trade-offs between different dimensions of the questionnaire, SF6D has an algorithm to create a continuous measure of health and was tested against a large sample of the UK population [REF-53].

There has been a proliferation of measures of quality of life and health status over the last two decades. In 2000, 1,275 separate measures existed, and the production of new measures was considerably growing [REF-54]. As such, there is a need for generic, comparable measures. Amongst this variety of measures, previous studies have highlighted three instruments important for measuring PROMs: the SF-36® [REF-52], PROMIS 10 [REF-39] [REF-55] and EQ-5D [REF-49]. All three instruments provide a means of describing (generating a "profile" of) health. The SF-36® applies an algorithm to patients' responses to individual questionnaire items and scales to produce two summary scores: one for physical health and one for mental health. The EQ-5D uses a visual analogue scale to elicit from patients a single score for their overall health. The PROMIS 10 does not yield an overall score but gives physical health and mental health component scores. Extensive research has been done on the validity, reliability, reproducibility, and utility of health status surveys when applied to general audiences and sub-groups based on age, sex, nationality, and disease entity. The EQ-5D and SF-36® have both been used in large surveys of the general public – population norms are available for both that may have relevance in "benchmarking" performance. Both generic measures – EQ-5D and SF-36® – have validated translations of their instruments available in a range of languages. The PROMIS instrument, being much newer, has a limited evidence base. Table 2 presents the comparison of the three most widely used generic PROM instruments.

Table 2: Comparison of the Three Generic PROMs Instruments.

PROM instrument	Domains measured	No. of questions	Discrimination between domains and pt. types and populations	Validation	Translation
SF-36	9 domains: Physical functioning (PF) (10), Rolelimitation-Physical (RP) (4), Bodily Pain (BP) (2), General Health (GH) (5), Vitality (VT) (4), Social Functioning (SF) (2), Rolelimitation-Emotional (RE) (3), Mental Health (MH) (5); health transition (1)	36	Able to detect differences between groups defined by age, sex, socio-economic status, geographical region and clinical conditions	Validated across a range of conditions, settings and languages	170 languages
EQ-5D	4 domains: Anxiety/Depression (1), Mobility (1),Pain/Discomfort (1), Self-Care (1), UsualActivities (1)	5	underperforms in the Pain/Discomfort dimension	Validated in a diverse patient population in several countries, including patient groups with chronic conditions (cardiovascular disease, respiratory disease, depression, diabetes, liver disease, personality disorders, arthritis, stroke)	170 languages
PROMIS Global 10	5 domains: physical function, fatigue, pain, emotional distress, and social health	10	N/A	Newer than the previous two measures. Validated for some conditions and patient groups	8 languages

Source: Measuring Generic Patient Reported Outcome Measures for Health System Improvement, OECD, 2017. OECD Workshop on Generic PROMs.

The SF-36 is a widely used generic measure of health status. Thirty five of the 36 items are grouped into eight scales that address health constructs considered to be important to most health care situations: physical functioning, role limitations (physical problems), bodily pain, general health, vitality, social functioning, role limitations (emotional problems), and mental health. One item assesses the perception of changes in health but is not used to compute scale scores. The SF-36v1® was developed in 1990 during the Medical Outcomes Study (MOS) to measure generic health concepts relevant across age, disease, and treatment groups. It is available in 161 languages [REF-52]. The EQ-5D has two components: EQ-5D Visual Analogue Score asks patients how good or bad [their] health is today, on a scale of 0 (worst) to 100 (best). The second component, the EQ-5D index score, asks patients to indicate their current health status in dimensions of mobility, ability to undertake self-care, ability to undertake usual activities, pain/discomfort, and anxiety/depression. It was developed in 1990 and is available in 170 languages. The PROMIS Global Health questionnaire was developed by the National Institutes of Health in 2009. The questionnaire includes a 10-question survey that assesses generic health-related quality of life compared with population norms. PROMIS 10 gives a summary indicator of health status by assessing five domains: physical function, fatigue, pain, emotional distress, and social health. Nine of 10 questions are answered using 5-point Likert scales, and the 10th question is answered using a numerical rating scale.

3.2.5 EQ-5D & Quality Adjusted Life Years (QALYs)

The PROM, EQ5D is a utility measure, permitting comparison of outcomes across diseases, similar to other utility measures, HUI and SF6D. EQ5D measures QoL which when combined with duration creates the measure, quality adjusted life years (QALYs). The QALY is a measure of the value of health outcomes. Since health is a function of length of life and quality of life, the QALY was developed as an attempt to combine the value of these attributes into a single index number. The QALY calculation is simple: the change in utility value induced by the treatment is multiplied by the duration of the treatment effect to provide the number of QALYs gained. QALYs can then be incorporated with medical costs to arrive at a final common denominator of cost/QALY. This parameter can be used to compare the cost-effectiveness of any treatment.

EQ5D exists as a number of forms e.g., EQ5D*, EQ5D-5L and EQ5D-3L (www.euroqol.org, accessed 1/11/2021). EQ5D-3L [REF-49] for example, has 3 levels of sensitivity for each of its questions, creating 3^5 or 243 unique health states for patients to choose between. When a value set was created for the UK, 43 states were valued by the public, using a technique known as time trade-off (“TTO”), from which the remaining states’ values were regressed [REF-50].

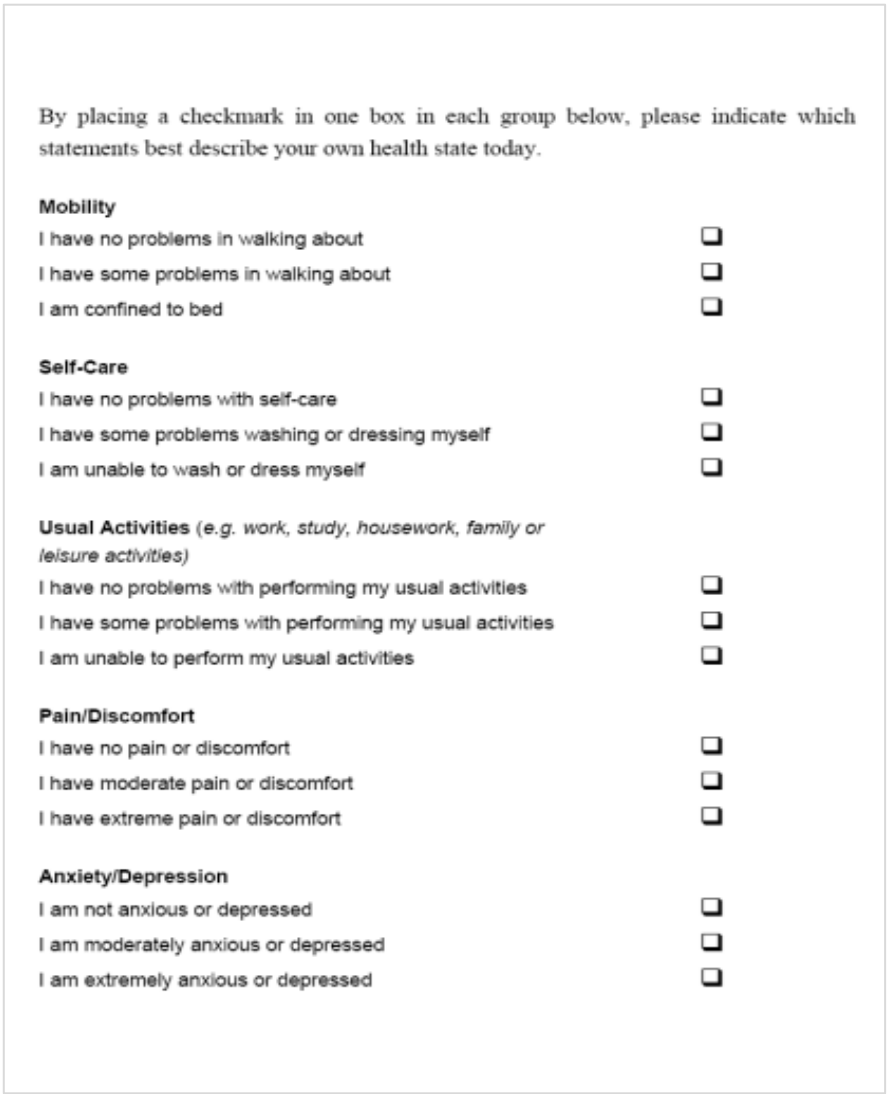
Use of the tool requests patients to answer a question, selecting a level of answer equivalent to their health at the time e.g., “…mark the option which best describes your health today” (for your) usual activities. There are 4 other question areas including mobility, self-care, pain/discomfort and anxiety/depression, creating 5 dimensions in total for which a patient selects from a choice of no problems, some problems or unable to perform for this question. The choices of response which patients are asked to select between are discrete i.e., non-overlapping, allowing a respondent to assign an answer definitively to one category of

response, helping create a reliable instrument, i.e., the same outcome, if used again at a different time, under the same conditions. This is a consequence of considerable rigour in PROM design, including use of patient input to determine matters of significance to patients and clarity of questions to stakeholders.

Individual EQ5D-3L index scores range between -0.594 (for the state 33333) and 1.000 (for the state 11111). The highest value is assigned to patients who report the best possible health state for each of the five domains. 3L values are used by the UK, Australian and Canadian health technology agencies, respectively National Institute for Health and Care Excellence (NICE), Pharmaceutical Benefits Advisory Committee (PBAC) and the Canadian Agency for Drugs and Technology in Health (CADTH) (Euroqol, accessed 30/9/2021).

EQ-5D is considered the PROM of choice in order to be able to measure QALYs and is the most common generic measure of the international literature in cost-utility analysis.

The EQ5D-3L questionnaire is presented in Figure 3.



By placing a checkmark in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about ❑
I have some problems in walking about ❑
I am confined to bed ❑

Self-Care

I have no problems with self-care ❑
I have some problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities*)

I have no problems with performing my usual activities ❑
I have some problems with performing my usual activities ❑
I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑
I have moderate pain or discomfort ❑
I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑
I am moderately anxious or depressed ❑
I am extremely anxious or depressed ❑

Figure 3: EQ-5D-3L Questionnaire.

Use of this questionnaire is subject to permission by Euroqol (www.euroqol.com).

EQ5D-3L is the recommended PROM for the HosmartAI study as it is validated, exists in many languages and is used by a number of health technology agencies. However, the choice of PROM is open to each study, who should select the PROM most valuable to their programme.

For the HosmartAI project which is being undertaken in a number of countries in the EU, a well-known PROM which is validated against locally measured populations is ideal. EQ5D-3L is a validated instrument, it is commonly used by health technology authorities and is available in many languages. It is also short and simple to use and self-filled. EQ5D-3L consists of 5 domains of question reflecting 3 levels of patient's perception of their health that day in those areas. Most pilots have been willing to consider using EQ5D-3L, although some have other/additional instruments that may be used e.g., Pilot 5, a national instrument and Pilot 6, SF36. Of the 5 questions EQ5D-3L poses, 3 levels of answer (none, some, much) are accorded greater decremental values the more severe the disablement. The sum of decrements subtracted from the maximum quality of life, weighted at unity i.e., 1 gives the QoL. The current range of tariffs are listed at EuroQol (https://euroqol.org/) [REF-49]. Thus, EQ5D-3L is recommended [REF-49].

3.3 Patient Reported Experiences Measures (PREMs)

Health systems are searching for ways to make their services more responsive to patients and the public. Often there is a perceived need to respond to consumer pressure and to make health care more like other consumer experiences, which has advantages and disadvantages that we are going to analyse further below. In order for a health system to be able to identify consumer/patients' experiences, there is a need for patient engagement. The patient is encouraged to take an active role as a key player in protecting their health, choosing appropriate treatments for episodes of ill health and managing chronic disease.

Considerable evidence suggests that patient engagement has been proved to improve patient's satisfaction with a positive impact on their clinical and economic aspects. The distinct role of patients in health care is presented in Table 3.

Table 3: The Distinct Role of Patients in their Health Care.

Patients can play a distinct role in their health care by:
• **understanding the causes of disease and the factors that influence health;**
• **self-diagnosing and treating minor self-limiting conditions;**
• **selecting the most appropriate treatment for acute conditions, in partnership with health professionals;**
• **managing treatments and taking medications appropriately;**
• **monitoring symptoms and the effects of treatment;**
• **being aware of safety issues and reporting them;**
• **learning to manage the symptoms of chronic disease; and**

- **adopting healthy behaviour, to prevent the occurrence or recurrence of disease.**

Recognizing their role and seeking to strengthen it is fundamental to securing a more patient-centred approach to health care delivery. It also provides the essential underpinning for strategies that aim to reduce health inequalities and improve health for all.

PREMs seek to determine how patients find the quality of their care from the patient's perspective. Similar to PROs, patient experience is also measured using surveys or questionnaires. These can be administered in various ways. A number of approaches and questions have been developed. Questions can be tailored to a certain setting e.g., primary care, home, hospital, long-term care or assess a specific aspect of care (e.g., continuity, autonomy, quality of information provision). PREMs, developed from a patient marking a 0-100 rating scale "thermometer" according to their satisfaction with the subject of the question e.g., quality of today's service, fulfilment of needs, etc, have gone on to generate more granular instruments which seek to deconstruct the aggregate ratings scale score or broaden the range of questions e.g., how patients have found service accessibility, communication, continuity and confidence. These data are now used to inform development of health services and treatments [REF-60].

Collecting and using patient-reported data a range of factors influence the outcomes of care as reported by patients, including behaviour, adherence, age and comorbidities. But readmission and mortality are subject to the same confounding variables. Like any outcome data that are used for benchmarking, confounders for patient-reported indicators should usually be adjusted in order to enable meaningful comparisons [REF-61]. All data, whether patient-reported or not, have limitations and should be interpreted with the necessary caution [REF-62].

PREMs include myriad specific condition/treatment/population instruments. Alternative to patient considerations, user reported experience measures ("UREMs") might be considered as a means to determine treatment preference from different stakeholder groups e.g., clinicians/nurses/other hospital staff/carer, to answer questions relating to how easy the use of new treatment strategy was, how easy it was to use aspect abc, would you recommend the use of treatment strategy a or b to colleagues. Once again, answers to such questions help inform the development of better instruments.

3.3.1 PREMs Evaluation Methods

PREMs come in a variety of forms depending on the objective, the disease area, the patient manifestation and the technology which is going to be assessed. In the simplest form, a PREM might be used to answer how good was your treatment or experience today? It could be answered by the user marking a line across a 0-100 visual analogue (thermometer-like) scale, or Visual Analogue Scale (VAS) to determine a user's satisfaction with the treatment, where 100 represents absolute satisfaction and no satisfaction, as illustrated in Figure 4 below. The patient is asked to indicate the level of satisfaction from the health care service he used by drawing a line through the 0-100 thermometer scale. The lower the mark, the less satisfied

patient and vice versa. A mark at the bottom indicates absolute dissatisfaction and at the top absolute satisfaction.

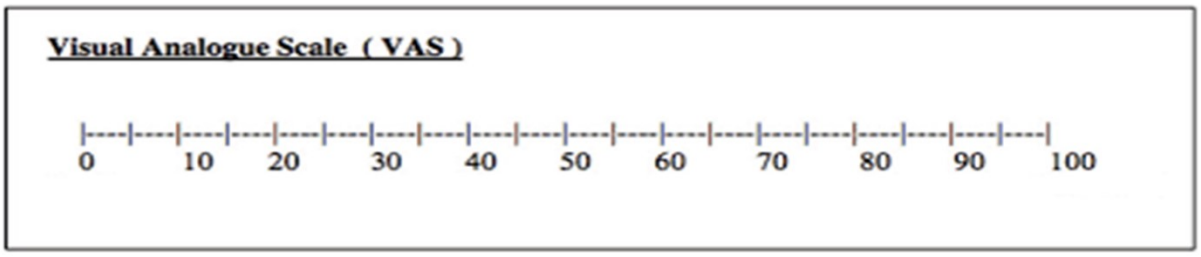


Figure 4: A Visual Analogue Scale for Assessing Patient Reported Experience of Treatment.

As mentioned in Section 3.2 regarding PROMs, a generic measure to measure Quality of life is Likert Scare, which is also used for patient experience measures, as described below.

Likert [REF-57] formed a mode analysis splitting progressively strengthening, ordinal categories of response to allow estimation of the frequency of responses over a continuum across responders through categorisation "bins" against questions, such as, how satisfied were you with today's service (or with x aspect of today's service, if drilling deeper) [REF-58][REF-59]. The bins, splitting a data series into not necessarily equally sized [REF-63], but non-overlapping bins to measure frequency over what should be discrete ranges e.g., for a 2-point scale on the question, how satisfied were you with the Likert scale explanation?

Not all **Fully**

The bins are clearly non-overlapping. As the number of bins increases, maintaining discrete distinction could become a challenge e.g., Over the last year, how often did you feel your clinical problems were completely met:

never, rarely, now and then, sometimes, frequently, always

where the difference between now and then and rarely or sometimes is debatable, which could create different responses and so unreliability if a study were repeated. Accordingly, deleting ambiguous bins i.e., "now and then" to produce fewer but more reliable and easily distinguishable interpretations would create a better instrument e.g., if you asked to rank from least to most frequent, the following terms, what would be the answer?

never, rarely, sometimes, frequently, always

In this case, rarely is easily understood as less frequent than sometimes and sometimes less frequent than frequently and never the least frequent.

Indeed, infinite division of categories approaches reversion to a continuum, where integration of the area under the curve would define a quantitative response value to a single question, but with less intuitive interpretability than a limited number of ordinal category bins exploring qualitative progression, defeating the purpose of a categorization.

Likert scales also be constructed to measure:

- Strength of agreement

- Frequency of occurrence
- Pre-event quality expectation
- Post-event quality ranking
- Importance
- Likelihood

VAS or Likert scales might be used to gather patient satisfaction for example where there is a low or mixed reading age [REF-57]. However single item scales have previously been reported as lacking reliability and so validity [REF-64]. Accordingly, more sophisticated instruments, disease or treatment specific should at least be considered. There are many disease-specific PREMs, which are similar to PROMs.

PREMs represent an alternative perspective of how well a treatment seemed today, from the patient or alternative user's (carer or clinician) perspective. Some examples of PREM questionnaires are presented with a short description in Table 4 below:

Table 4: Patient Reported Experience Questionnaires.

Patient Reported Experience Questionnaires	Reference
PSQ-18 -The Patient Satisfaction Questionnaire Short Form a Likert scale patient satisfaction questionnaire short form (PSQ-18) uses 7 domains of question, namely general satisfaction, technical quality, interpersonal manner, communication, financial aspects, time spent with doctor and accessibility and convenience. [REF-65]	Thayaparan, A. J., & Mahdi, E. (2013). The Patient Satisfaction Questionnaire Short Form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. Med educ online, 18, 21747. https://doi.org/10.3402/meo.v18i0.21747
ENDOPREM: Patient Experience of GI Endoscopy Questionnaire in both Routine Clinical Care and Research Studies:5 Domains - Anxiety, Expectations, Choice & control, Communication & Information, Embarrassment & Dignity. [REF-66]	Neilson LJ, Patterson J, von Wagner C, et al. (2020) Frontline Gastroenterology 11:209–217.doi:10.1136/flgastro-2019-101321
PSNCQQ - Patient Satisfaction with Nursing Care Quality Questionnaire: 12 Domains - nursing and daily care, ancillary staff and hospital environment, medical care, information, admissions, discharge and billing, overall quality of care and services, recommendations and intentions, overall quality of care during the hospital stay, overall quality of nursing care, intention to recommend the hospital to family and friends and overall health outcomes. [REF-67]	Laschinger, H.S, McGillis L, Pedersen, C; Almost J. (2005) A Psychometric Analysis of the Patient Satisfaction With Nursing Care Quality Questionnaire, J Nurs Care Quality 20(3):220-230.

Alternative to patient considerations, User Reported Experience Measures (UREMs) might be considered as a means to determine treatment preference from different stakeholder groups e.g., patient/ clinicians/ carer. There appears to be no consensus on a particular PREM or group of PREMs as dominating use. Accordingly, pilot leaders must make their own choices or whether they need to address user or patient satisfaction or experience.

A few examples of User Satisfaction Experience Measures are described in Table 5.

Table 5: User Related Experience Questionnaires.

Questionnaire & Domains	Reference
SUTAQ – Service User Technology Acceptability Questionnaire Digital system assessment by end users Questionnaire Domains: Perceived benefit, privacy and discomfort, care personnel concerns, Kit as substitution and satisfaction. [REF-68]	Torbjørnsen, A., Småstuen, M. C., Jenum, A. K., Årsand, E., & Ribu, L. (2018). The Service User Technology Acceptability Questionnaire: Psychometric Evaluation of the Norwegian Version. JMIR human factors, 5(4), e10255. https://doi.org/10.2196/10255
SUS - System Usability Scale Ease of use or functional operation of technology (or lack thereof) software, hardware, mobile devices, and other technological applications. Questionnaire Domains: Effectiveness, efficiency and satisfaction.[REF-69]	Brook J. (2013) SUS: A Retrospective. J Usability Studies 8(2):29-40 JUS Journal Template (uxpajournal.org)
UEQ - User Experience Questionnaire Questionnaire domains: attractiveness - perspicuity - efficiency – dependability – stimulation and novelty using a new technology. [REF-70]	Hinderks A., Schrepp M, Domínguez Mayo MJ, Escalona MJ, Thomaschewski J (2019), Developing a UX KPI based on the user experience questionnaire, Comp Stand Interfaces, 65:38-44, https://doi.org/10.1016/j.csi.2019.01.007.

3.3.2 Policy Makers & Patient Reported Experiences Measures (PREMs): The case of OECD PaRIS initiative

In 2017 the OECD Health Committee launched the Patient Reported Indicator Survey "PaRIS" initiative with the objective of health systems to become more people-centred by developing international benchmarks of health system performance based on data reported by patients themselves [REF-70]. Health systems are in need of better information about the value and outcomes they produce. In this committee it was identified that there is little information available about the impact of health care services upon the people served, beyond re-admissions to hospital, complications and deaths, thus there was a strong need to assess health care outcomes from the perspective of the people served. PaRIS was initiated in order to build international capacity to measure and compare patient-reported indicators, using indicators that enable comparisons across countries. It also aimed to encourage patient-reported indicators to evolve in a common direction internationally, to enable shared learning, development and research. This is key to learning how well health services deliver their ultimate objective: supporting people in regaining and sustaining their health and well-being [REF-71].

Given the global trend of increased expenditure on health care as a share of national income, it is surprising that systematic, empirical measurement of the outcomes and experiences of care from the patient's perspective is still the exception in most health care systems. This gap in knowledge limits the ability for evidence-based policy making and the ability to maximise the benefits of health care at acceptable costs. It is difficult to improve what is not being measured. The PaRIS initiative addressed this knowledge gap.

The general objective of the PaRIS initiative was to develop, pilot and implement new patient-reported indicators of health system performance, specifically patient-reported experience measures (PREMs) and patient-reported outcome measures (PROMs). PaRIS contributed to health systems to become more people-centred by providing systematic, internationally standardized information on what matters most to patients.

The most important PREMs for the HosmartAI technologies are patient/physician satisfaction and user experience measures. Both metrics will provide important insights into the patient-centric domain of the AI technologies.

3.4 Hospital Efficiency/Productivity Measures

Health care efficiency is a growing issue in most European countries where health care expenditure is rapidly increasing. Therefore, accurate productivity metrics are essential to avoid sub-optimization within the health care system. A major feature of health-care systems is substantial variation in hospital productivity. Hospital productivity varies widely across countries. The presence of such variation suggests potential areas for improvement, which can substantially lower health care costs.

This trend has resulted in the proliferation of quality indicators to measure performance and outcomes. Indicators may be used to monitor the quality of care in a single institution or across the health care system, to promote quality improvement activities, to make comparisons over time between institutions (Benchmarking), or to assist consumers to choose health-care providers [REF-72].

Quality measurement can be a costly and time-consuming activity, and thus the judicious selection of indicators that contribute to an aggregate understanding of health-care quality is imperative. We have argued previously that indicator development should proceed in a systematic fashion, targeting areas where the need is greatest, and have described a framework to assist with this aim. For this to be achieved, it is first necessary to identify and classify areas of need. Thus, the effort posed was to identify and classify clinical and process indicators currently being used in various countries to measure the quality of care provided by hospitals, and to identify commonalities in measurements in order to be able to apply at HosmartAI technologies. Specifically, we aimed to identify and classify indicators according to:

- Domain to which they apply (hospital-wide, surgical and non-surgical clinical specialties and departments).
- Aspects of care provision (structure, process outcome).

- Dimensions of quality (safety, effectiveness, efficiency, timeliness, patient-centredness and equity). This information was then used to identify gaps in current measurement.

For health production processes of any complexity, there are usually a number of stages in the transformation of resources to outcomes, and much of the confusion in discussing efficiency arises because commentators are discussing different parts of that process. To illustrate, Figure 5 represents a typical (but simplified) process associated with the treatment of hospital patients. The overarching concern is with cost-effectiveness, which summarizes the transformation of costs (on the left-hand side) into valued health outcomes (the right-hand side). However, the data demands of a full system cost–effectiveness analysis are often prohibitive, and the results of such endeavours may in any case not provide policymakers with relevant information on the causes of inefficiency, or where to make improvements. To take remedial action, decision-makers require more detailed diagnostic indicators of just part of the transformation process.

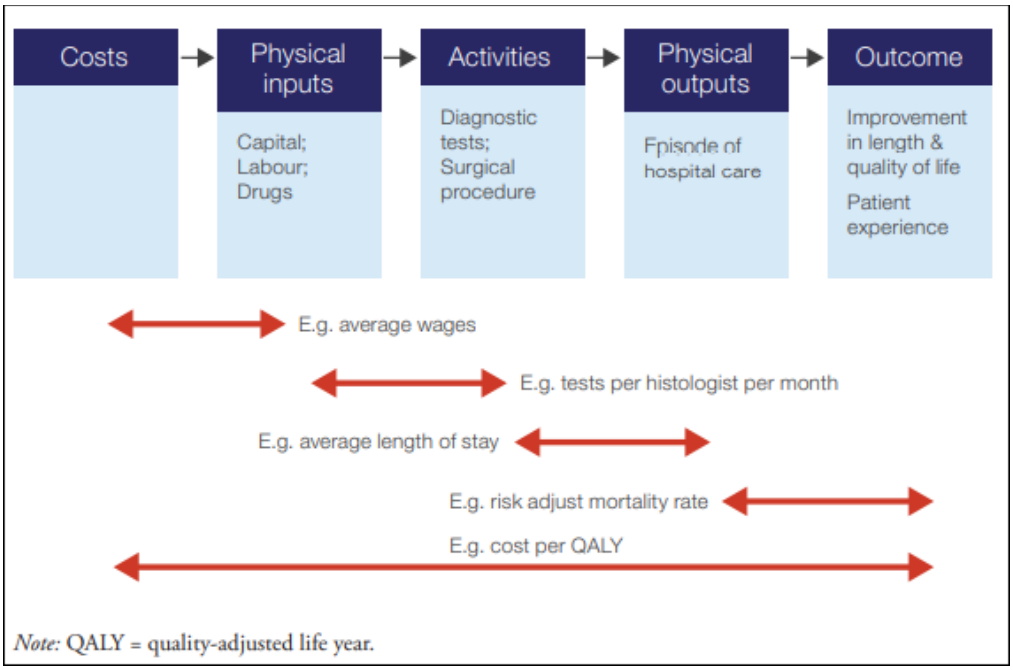


Figure 5: The Production Process in Hospital Care.

Physical outputs are created by aggregating activities for a particular service user. In a hospital setting, this usually refers to single episodes of patient care, an aggregation of many actions such as tests, procedures, nursing care and physician consultations. There is great scope for waste in this process, for example, in the form of duplicate or unnecessary diagnostic tests, use of branded rather than generic drugs, or unnecessarily long length of stay. Much depends on how the internal processes of the hospital are organized so as to maximize outputs using the given inputs. The well-known metric of length of stay, which indicates the number of bed days expended per case, falls into this category. The final stage of the health system production process is the quality of the outputs produced. Even when they employ the same physical inputs, activities or physical outputs, there is great scope for variation in effectiveness

among providers. The notion of quality in health care has a number of connotations, including the clinical outcomes achieved (usually measured in terms of the gain in the length and quality of life) and the patient experience (a multidimensional concept). So, for example, even though two hospitals produce identical numbers of hip replacements, because of variations in clinical practice and competence, the value they confer on patients (in the form of length and quality of life, and patient experience) can vary considerably. Quality-adjusted output is usually referred to as the outcome of care in the literature. Quality of care has become a central concern of policymakers, and its measurement, while contentious, is usually essential if a comprehensive picture of efficiency is to be secured. Note that the unit costs metric usually links costs to physical outputs. The numerous partial efficiency indicators that have been developed seek to shed some light on the reasons for variations in unit costs. Some metrics, such as the labour productivity or length of stay examples, are based on only partial measures of inputs or outputs. Some are capable of adjustment for external influences on attainment, others are not. None addresses the production process in its entirety, that is, the cost–effectiveness with which costly inputs are converted into valued outputs.

Furthermore, this example looks only at the hospital sector, without reference to other aspects of the health system. It, therefore, focuses mainly on hospital productivity, making no judgement on other issues, such as whether patients might have been treated more cost-effectively in different settings (for example, primary care or nursing homes). And by focusing on the curative sector, it can shed no light on the success or otherwise of the health system's efforts to prevent or delay the onset of disease. A further aspect of whole system performance that is ignored is the impact of hospital performance on other sectors within the health system. For example, it may be the case that apparently high levels of efficiency in the average length of stay are being secured at the expense of heavy workloads for rehabilitative and primary care services, which may or may not be efficient from a whole system perspective.

HosmartAI pilots' will define, gather and analyse their own measures of productivity/process outcomes, which vary considerably according to the nature of the pilot.

Each organization, like a hospital or health centre, consumes a series of physical resources, referred to as inputs, often measured in terms of total costs. The organization then transforms those inputs into a series of valued outputs. Although measuring the aggregate value of inputs in terms of total costs is relatively uncontroversial, the valuation of aggregate outputs in the health sector depends on how much importance we place on different health system outputs, such as health improvement and quality of life. Nevertheless, if we can agree on a measure of aggregate valued outputs, then we can calculate a summary measure of efficiency as the ratio of valued outputs to inputs, what is often referred to as cost–effectiveness, or how well the organization's costs are converted into valued benefits.

Almost all efficiency analysis relies on comparisons, similar to HosmartAI, so it is important to ensure that the entities being compared are similar. A great deal of efficiency analysis is concerned with securing such comparability. If organizational entities are operating in different circumstances, perhaps because the population cared for or the patients being

treated differ markedly, some sort of adjustment will be needed to ensure like is being compared with like.

3.4.1 The Joint Learning Network for Universal Health Coverage – World Bank

In 2020 an initiative from World Bank was performed in order to produce a Resource Guide by the Joint Learning Network for Universal Health Coverage (JLN), an innovative learning platform where practitioners and policy makers from around the globe co-develop global knowledge that focuses on the practical "how-to" of achieving universal health coverage [REF-72].

The Joint Learning Network (JLN) is a global network connecting practitioners and policy makers from 31 countries around the globe. This initiative decided on two strategic work streams: i) the measurement and information stream (MIS) which aimed to provide a framework for identifying and measuring efficiency, and ii) the systematic priority setting stream (SPS) which aimed to support countries in maximizing their stated health sector priorities within a given resource envelope [REF-72].

The JLN identified a list of inefficiencies in health which are categorized into two types:

1. Inefficiencies arising from system-level resource allocation decisions – from poor planning or slow response to the changing health needs of the population:
    - Inappropriate or costly input/staff mix.
    - Inappropriate hospital size.
    - Sub-optimal deployment of health workers and facilities.
2. Inefficiencies that result from facility- or physician-level decisions linked to poor incentives or lack of accountability measures:
    - Inappropriate hospital admissions or length of stay.
    - Over-use of health care technology.
    - Sub-optimal quality of care and medical error.
    - Under-use of generic drugs.
    - Irrational use of drugs.

Several characteristics of the health sector make health spending particularly prone to inefficiency. Uncertainty in the demand for health, informational asymmetries between patients and providers, difficulties in linking inputs to outcomes, and fragmented sources of financing often lead to lower actual spending on health than the allocated budget.

The guide which was produced by this initiative provided a brief overview of concepts and principles of efficiency, and a framework for identifying and measuring efficiency in a practical way. It provided a list of indicators most often used for tracking health system performance and gives guidance on how they can be used to measure efficiency [REF-72] [REF-73].

Efficiency analysis attempts to explain the unexplained variation across accountable entities – that is why some individual providers, facilities, or health systems perform better than others and benchmarking makes a comparison based on average performance, relative to the best performer, relative to a clinical norm or target, or relative to past performance.

In the case of HosmartAI, all AI technologies will be compared with the standard of care or currently used technology in order to be able to identify the added value of the new technology.

Based on JLN the most basic definition of efficiency is maximizing outcomes relative to inputs.

- Inputs: Most often, people use inputs to refer to the costs, resources, or investments used to buy or produce health care inputs; and while financing for the health care system is one of the most fundamental inputs, others include health workforce (e.g., doctors, nurses, midwives, community extension workers); physical infrastructure (e.g., health care facilities, medical supply stores); drugs and medical products, equipment (e.g., MRI machines), and information – often the most overlooked input (e.g., data on civil registration and vital statistics, disease-specific registries, patient reported health outcomes, drug stocks).
- Outcomes can refer to the consequences, effectiveness, or benefits of service delivery interventions. In most health care systems around the world, the main outcomes of interest concern health status, financial risk protection, and public satisfaction. However, in practice, most efficiency metrics use intermediate outcomes or outputs. Outputs include information on the quantity (e.g., availability, access, coverage) and quality (e.g., diagnostic accuracy, treatment success rates) of the goods and services provided.

The JLN team produced a list of indicators to routinely assess the efficiency of health sector spending, especially in areas that are known to be major sources of inefficiency and consume the most resources, such as hospitals and pharmaceuticals (Table 6)[REF-72].

Table 6: Common Indicators used to assess efficiency in pharmaceutical and hospital sub-systems.

Pharmaceuticals	Hospitals
• Pharmaceutical spending as % of total health expenditure (THE) • Antibiotics spending as % of total pharmaceutical spending • Unit price of drugs/medical consumables • Unit price compared to international reference prices (especially for high-cost/use items) • Cost of freight/distribution to facilities • Order/use of high-cost items • High-use items • Number or % of expired items • Value of expired items • Stock-outs • Antibiotic prescription rates • Percent of encounters that end up in antibiotics being prescribed • Time to process orders • Time to pay suppliers • Drug availability • Rate of anti-microbial resistance	• Spending by function (e.g., outpatient, inpatient, pharmaceutical, primary health care, public health or prevention, curative care) as % of government health expenditure (GGHE) • Hospitals per 100,000 population, hospital bed density, bed occupancy rate • General service readiness • Number of visits/admissions per day/month/year/per capita • Share of outpatient/inpatient • Diagnostic accuracy for tracer condition • Adherence to clinical guidelines • Number of incidents per 1,000 patient days (e.g., center line-associated bloodstream infections, standardized infection ratio) • Avoidable admissions for chronic obstructive pulmonary disease (COPD), asthma, hypertension, diabetes • Referral rate • Average length of stay • Readmission rate • Caesarean section (C-section) rates

The final list of indicators presented eventually used reflected a balance of what was available at the time of the research in participating countries as well as some aspirational indicators. There were many health indicators lists that had been developed by international organizations, academics, advocacy groups, and others (including the JLN) grouped for different purposes – for tracking progress towards universal health coverage (UHC), assessing primary health care (PHC), ensuring hospital quality, and/or evaluating health provider payment systems. The final list of indicators which was eventually used and analysed is presented in Table 7 [REF-72].

Table 7: Common Indicators that can be used in efficiency analysis – JLN.

Financing (management)	Inputs (management)	Outputs (access)	Outputs (quality)	Outputs (risk factors)	Outputs (management)	Outcomes (health status)	Outcomes (financial protection)
• Total health expenditure (THE) as % of GDP or in per capita terms • Government health expenditure (GGHE) as % of GDP, as % of budget, as % of THE, or in per capita terms • Share of pre-paid/pooled spending as % of THE; out-of-pocket expenditure (OOP) as % of THE; external funds as % of THE • THE, GGHE, OOP spending by function (e.g. outpatient, inpatient, pharmaceutical, primary health care, public health or prevention, curative care) as % of THE, as % of GGHE, as % of OOP expenditure, or in per capita terms	• Density of doctors, nurses, midwives • Hospital density or hospital bed density • Availability of basic facility infrastructure, essential medicines, equipment • General services readiness or service-specific readiness for tracer condition(s)	• Service utilization • Antenatal coverage • Births attended by skilled health personnel • Immunization coverage • TB case detection	• TB treatment success rate • Diagnostic accuracy for tracer condition(s) • Adherence to clinical guidelines for tracer condition(s) • Diabetes control • % of incidents per 1,000 patient days • Avoidable admissions	• Raised blood glucose/diabetes among adults • Children under 5 who are stunted • Tobacco use among persons aged 18+	• Caseload • Absenteeism • TB notification rate • Average length of stay in hospital • Bed occupancy rate • Referral rate • Hospital readmission rates • C-section rates • Expired drugs • Stock-outs • Claims ratio • Budget execution rates	• Life expectancy at birth • Under-five mortality rate • Maternal mortality rate • TB burden • Mortality between 30–70 years age from cardiovascular diseases, cancer, diabetes or chronic respiratory diseases	• Catastrophic health expenditure

Note: Based on stated country priorities, indicators for immunization and maternal health were used as tracer conditions for assessing efficiency of primary health care, tuberculosis as a tracer condition for assessing efficiency of communicable diseases and a functioning surveillance system, and diabetes as a tracer condition for assessing efficiency of non-communicable diseases. Maternal health and diabetes can also be used as proxy conditions for assessing the efficiency of a functioning referral or integrated system.

The challenges and lessons learnt from the JLN project which are presented below, were thoroughly examined by the research team of HosmartAI in order to avoid similar drawbacks.

Challenges:

a) Availability of data: A key limiting factor was the availability of data at the right level of disaggregation to make meaningful comparisons (e.g., by geographical region, across population groups) – especially for decentralized countries. It was also difficult to get data across different parts of the health system as information is held in different departments or programs within the Ministry of Health or was spread out across several Ministries/Agencies or levels of government. However, when pressed, all pilot countries were able to bring in anecdotal evidence or refer to country studies to help explain their data.

b) Defining the scope of the efficiency analysis: All four countries attempted to present country data for the entire list of priority indicators. This made it harder to focus on any one area and to delve deeper into identifying where the results chain inefficiencies might lie.
c) Interpreting the data: There was little interpretation of what the data might mean and/or packaging the information into a policy-relevant storyline [REF-72].

Lessons learned:

a) Focus the scope of the analysis: That is, looking at hospital efficiency, pharmaceuticals or a specific tracer condition within primary health care, such as maternal health or tuberculosis, where relevant indicators were more likely to be available under responsible units or budget holders.
b) Additional indicators should be analysed when specific issues emerge in the course of routine monitoring: This includes going beyond indicators and drawing on practitioner's valuable knowledge of the sector to help fill in the gaps in the storyline when data is not available. This will help with the interpretation of findings and the assessment of options for further action.
c) Efficiency analysis needs to be embedded into the formal decision-making process, not a one-off exercise. Many health system performance indicators are already regularly collected and reported as part of the routine monitoring and evaluation of the health sector. Building on existing processes and applying an efficiency lens to reviewing indicators that question why some regions, hospitals, or providers perform better than others is the first step to institutionalizing efficiency analysis [REF-72].

3.4.2 The EuroHOPE project

At the European level, the great diversification of indicators was identified and the European Commission funded in 2013 the EuroHOPE project, the construction of an international comparative database which allowed performance indicators to be calculated at national, regional and hospital levels for several different disease groups [REF-73]. In the EuroHOPE project, the performance indicators were developed in collaboration with clinical experts in the different disease groups, and with experts in health economics, epidemiology and statistics. The disease-based approach required patient-level data covering the whole population and the possibility to deterministically link records from different national registers. In the seven countries (Finland, Hungary, Italy, the Netherlands, Norway, Scotland and Sweden) included in the EuroHOPE project, it was possible to link national hospital discharge registers with mortality registers and with registers of medicines prescribed. In Italy, similar data were available for two geographical areas. All databases present population data that reflect patterns of care and outcomes for the entire population residing in the defined territories.

The EuroHOPE project followed the ideas of the PERFECT project so that database creation was based on several general stages: 1) define the patient population; 2) collect the register material for the relevant patient population; 3) define the start and end of the episode (by defining and using the index admission and deciding how referrals should be treated) for the

patients from the available variables concerning the care given; 4) check the history and follow-up on the use of health care services to define state and time variables for each patient; 5) construct the comorbidity variables; 6) calculate the direct health care costs; and finally, 7) combine the information from the previous stages to generate the comparison database. This was very challenging in an international context because of variations across data sources and differences in health system structures and practices. It was extremely difficult to find compromises that worked in each country and allowed for cross-country comparability. Also, in the estimation of the risk adjustment models, even after the standardized definition and data collection, a complication arose from the involvement of many different countries. Ideally, the individual-level data from all participating countries should be pooled before estimating the risk adjustment models, but that is not feasible because not all countries allow individual-level data to be shared because of privacy regulations. To avoid such problems, parameter estimates for the confounding factors were first estimated for every process or outcome measure using registry data to compare health care efficiency using the broadest possible pooled data for each disease. Then, the coefficients of each model were made available to all partners who then calculated individual level-predicted values for the indicators. The predicted values were then summed up at the country and regional level. The ratio of observed and the predicted value of the dependent variable in the comparable unit could be multiplied by the average value of the indicator in the pooled data to calculate the risk-adjusted indicator [REF-74]. In practice, after definitions had been agreed for the required standard form of comparison data, each national partner was individually responsible for producing its own national EuroHOPE comparison data, with the principles stated in the disease-specific study protocols [REF-74]. After this, the partners used a common statistical code which automatically processed the data, extracted the coefficients for the models from the EuroHOPE server and calculated the predicted and risk-adjusted values at all levels. Finally, the descriptive statistics along with the country-, regional- and hospital-level indicators and their confidence intervals were automatically transferred to a reporting template. For five countries, it was possible to pool individual-level data and, using a more sophisticated methodology (multilevel modelling), analyse the hospital-level variation in 30-day survival and the cost of the first hospital episodes, as well as the relationship between the measures, that is, the existence of cost–quality trade-off [REF-75]. Generally, the study did not find a positive correlation in the pooled analysis and in the separate country-level analysis. The only exception was Sweden where an increase of cost from €5,000 to €20,000 was associated with an increase in 30-day survival from 90% to almost 100%. Further research could assess whether such spending increases provided good value for money.

The findings and analysis from the EuroHOPE project were very insightful for the HosmartAI efficiency measures (KPI) due to the common challenging issue of both projects, common foundation of different partners, countries, technologies and disease areas. That was the reason, at least at the early stage, that the research team decided to proceed with a separate analysis per AI technology/partner rather than the final platform [REF-76].

3.5 Economic Outcomes

Economics assesses the optimal use of resources to produce beneficial outcome for society. Cost analysis for various treatments is a useful component of outcome analysis to allow the weighting of other outcomes against unit cost or the inverse, cost per unit effect [REF-77]. Costs may be categorized as capital (amortized or one-off) and resources which are recurrent costs. While all costs will need a unit cost record by tier, recurrent costs will also need a volume per unit time. Accordingly, economic evaluation of health care treatment options assesses among other outcomes, cost, a common component of health care economic evaluations. Other outcomes include clinical effectiveness and health economic outcomes such as PROMs including utility measures such as "QALYs". As demand typically exceeds supply for health care, health care suppliers seek value for money. Accordingly, economic evaluations use opportunity cost to identify the best value option out of a choice of the intervention and the next best alternative treatment, the comparator, typically usual care, the current treatment [REF-78].

3.5.1 Combining Cost & Effect

In health economics, the cost alone does not translate the value of a specific treatment or technology without considering the effect of the specific technology. The combination of the lower cost and higher beneficial effect outcomes is the desired composition for the decision-making process. Yet, if each outcome falls on a different treatment option, a determination is still needed between options, with a consistent approach i.e., without reverting to personnel preference which can create a varied health care accessibility landscape for treatments across a country. The answer lies in using a ratio of incremental cost divided by incremental effect where the intervention's outcomes are used after subtraction of the comparator's equivalent outcomes e.g., intervention cost-comparator cost divided by intervention effect minus comparator effect [REF-46]. This creates the incremental cost-effectiveness ratio ("ICER") which may be written as indicated in Figure 6:

$$ICER = \frac{\text{Cost of intervention – cost of comparator}}{\text{Effect of intervention – effect of comparator}}$$

Figure 6: Incremental Cost Effectiveness Ratio, ICER.

where the effect may be expressed in clinical effect or health economic terms i.e., PROMs such as quality-adjusted life years (QALYs) or more typically, both.

The economic analysis should provide an estimate of the treatment strategies costs over different time periods e.g., over the course of the study, per treatment event and per annum. The economic analysis should be an estimate of the regular day to day operational cost for treatment of a patient with both treatment strategies and the incremental difference between the two (intervention minus comparator over the course of the study, with duration of each treatment stated).

For economic outcome, an analysis of direct medical costs, direct non-medical costs, indirect non-medical costs, symptom impact & adverse event costs should be measured per patient per unit time, with the unit being the duration of each treatment strategy. One-off and recurrent costs should be assessed for unit cost and resource quantity. Cost can be estimated from broadly 2 perspectives.

- The payer's or National Health System or
- The patient's or Society's perspective.

For the later, one must add the cost of carers unremunerated time and the patient's time off when sick [REF-46].

For the analysis of HosmartAI, the NHS provider perspective will be adopted. The NHS provider perspective includes treatment costs such as medicine costs, administration and monitoring, other health service resource use costs associated with the managing the disease (e.g., GP visits, hospital admissions), and costs of managing adverse events caused by treatment. It does not include patients' costs of obtaining care such as transportation, over-the-counter purchases, co-payments or time off work. Yet, it is up to technology pilots if they wish to gather more data e.g., on productivity losses arising from patients' inability to work, charged according to their policy at either a common minimum rate or the national average wage in order to be able to present the social perspective.

The cost to health payer whether government department or insurer for each treatment strategy equates to:

Departmental Cost = n patients * mean patient cost per treatment strategy

Costs are separated into three categories which are further analysed below: a) direct medical costs, b) direct non-medical costs, c) indirect non-medical costs.

- Direct Medical Costs can include hospitalization (short- and long-term), outpatient follow-up, residential and day care, pharmaceutical interventions, laboratory testing. Costs of treatment adverse events should also be noted by severity.
- Direct non-medical costs include transport costs to and from hospital e.g., non-emergency ambulance for non-motile patients and paid carer giver time.
- Indirect non-medical costs include patient and unpaid carer time off work, charged either at a country's minimum wage or social security payments.

And likewise, the quality-of-life profile and QALYs generated from each treatment over the course of the study should be noted [REF-46].

For the economic analysis of HOSMARTAI technologies, an analysis of direct medical costs, direct non-medical costs, indirect non-medical costs, symptom impact & adverse event costs will be performed. The resource utilization should be measured per patient per unit time, following the micro-costing methodology, including the number and type of major resources of the patient needs. The product of unit cost and volume is used to determine the overall cost of each treatment strategy, noting recurrent costs quantity. Incremental outcomes from intervention and comparator may be calculated for clinical, health economic and economic outcomes whereby economic incremental outcome would be the sum of costs across treatment duration for each treatment strategy with comparator subtracted from intervention's cost. The equations are presented below:

Incremental Cost = Cost of intervention (new AI technology) – cost of comparator (current technology)

Incremental Effect = Effect of intervention (new AI technology) – Effect of comparator (current technology)

3.5.2 Types of Economic Evaluation

Several types of evaluation exist which can be distinguished by the conditions in which they are used.

- Cost-minimisation analysis is used when there is evidence that clinical outcomes of treatment strategies are the same. A treatment strategy may be chosen by whichever treatment has the lowest cost [REF-46] [REF-77][REF-78][REF-79].
- Cost-effectiveness analysis (CEA) - used when clinical outcomes quantity may vary between treatment strategies but are expressed in natural units of clinical effectiveness e.g., reduced mortality or morbidity [REF-78].
- Cost-utility analysis (CUA) is a form a form of CEA used when clinical outcome quantity may vary between treatment strategies, but health economic outcomes are expressed in utility scales i.e., quality of life, typically the patient's, allows comparison across sites, treatment strategies, conditions. Normally QALYs are used [REF-46][REF-77][REF-79].
- Cost-benefit analysis – a comparison of costs and benefits with both being expressed in monetary units for services received [REF-46][REF-77][REF-79].

- Cost consequence analysis is a form of CEA which presents outcomes in discrete categories without aggregation or weighting [REF-77].

3.5.3 Cost Consequence Analysis

Due to divergence of patient conditions and interventions, the research team decided to proceed with a cost consequence analysis per pilot as like for like comparison was not possible.

Cost-consequences analysis (CCA) is a form of economic evaluation where disaggregated costs and a range of outcomes are presented to allow readers to form their own opinion on relevance and relative importance to their decision-making context [REF-77]. This is usually done using a descriptive table to present the effectiveness results (primary and secondary outcomes) in a disaggregated format, together with the estimates of the mean costs with appropriate measures of dispersion associated with each intervention. The aim of the study determines the construction and assumption of any analysis.

CCAs have been recommended for complex interventions that have multiple effects [REF-77], and public health interventions which have an array of health and non-health benefits that are difficult to measure in a common unit. CCAs are not restricted to any viewpoint and so readers and decision makers can see the impact of their decisions in the whole spectrum of costs and outcomes. CCAs have been recommended for complex interventions that have multiple effects, for example, lifestyle education in diabetes [REF-77] and public health interventions which have an array of health and non-health benefits that are difficult to measure in a common unit [REF-78]. CCAs are not restricted to any viewpoint and so readers and decision makers can see the impact of their decisions on patient costs or on other sectors such as criminal justice [REF-79]. Similarly, outcomes are not restricted to health outcomes such as QALYs and can include other measures of wellbeing such as patient, or indeed staff, satisfaction. These non-health considerations are becoming increasingly relevant to NHS decision makers. CCA may be of particular value to funders that are more concerned with patient-orientated outcomes and intervention costs such as Charities and some NIHR research programmes, particularly those with less focus on final stage randomised control trials. CCAs may also be particularly useful in feasibility or pilot studies when it is not clear which costs and outcomes will be most relevant to future definitive trials. Given the limited funding available for feasibility studies and the scarcity of health economists, CCA can provide a less resource-intensive alternative if interventions have important economic consequences or a full comparative analysis is premature, but still provide an opportunity to pilot instruments used to collect economic data such as resource use and health-related quality of life [REF-78]. Similarly, outcomes are not restricted to health outcomes such as QALYs and can include other measures of wellbeing such as patient, or indeed staff, satisfaction. These non-health considerations are becoming increasingly relevant to NHS decision makers. CCA may be of particular value to funders that are more concerned with patient-orientated outcomes and intervention costs such as Charities and some NIHR research programmes, particularly those with less focus on final stage randomised control trials. CCAs may also be particularly useful in feasibility or pilot studies when it is not clear which costs and outcomes will be most relevant to future definitive trials. CCA can provide a less resource-intensive

alternative if interventions have important economic consequences or a full comparative analysis is premature, but still provide an opportunity to pilot instruments used to collect economic data such as resource use and health-related quality of life.

The CCA approach helps to refine economic methods, identify relevant costs and outcomes and generate hypotheses for definitive cost-effectiveness studies and perhaps most importantly, provides a broader and richer source of economic information increasingly needed by NHS decision makers. It provides a straightforward way to present cost and outcome data alongside each other for a new health technology and its comparator(s) in situations where complexity in the research design might otherwise be pervasive. An example would be comparing the costs and consequences of different models of care across a care pathway in an observational study. Given the methodological issues associated with this design, an initial CCA can provide initial information on where further focus might be beneficial.

The CCA approach helps to refine economic methods, identify relevant costs and outcomes and generate hypotheses for definitive cost-effectiveness studies and perhaps most importantly, provides a broader and richer source of economic information increasingly needed by NHS decision makers. It provides a straightforward way to present cost and outcome data alongside each other for a new health technology and its comparator(s) in situations where complexity in the research design might otherwise be pervasive. An example would be comparing the costs and consequences of different models of care across a care pathway in an observational study. Given the methodological issues associated with this design, an initial CCA can provide initial information on where further focus might be beneficial.

Models describing the disease aetiology/care pathway route through which a patient will transition as discrete states to allow costs and effects to be appropriately assigned and so through quantifying the effect of introducing a new technology into current health care pathways and routine health and social care system use. Model time horizons for accrual of effects and costs should be stated.

Health and social care system and personal social services costs resulting from or associated with the use of the intervention should also include acquisition (including infrastructure) and maintenance costs.

Cost consequence analysis provides a comprehensive presentation of the cost and value of the intervention of scope. It is a listing of all the relevant costs and outcomes or consequences of the interventions and may include the following components:

- Direct Medical costs
- Direct non-medical costs
- Indirect costs (time costs, productivity costs)
- Health-related quality of life impact
- Utility impact
- Clinical outcomes (including adverse events)

The ideal cost-consequence analysis would include all possible health outcomes or consequences in order to allow decision makers the ability to determine the intervention's likely impact on their budgets and on the health of their patients.

In Table 8 the basic categories of costs and effects for a Cost-Consequence Analysis are presented [REF-79].

Table 8: Cost-Consequence Analysis Table – Main Categories.

Perspective: Funder or social	Intervention A			Intervention B		Difference	
COSTS	Units	Costs		Units	Costs	Costs Mean (95% CI)	
Direct Medical Costs							
Intervention A/B							
Other medication/interventions							
Physician office visits							
ER visits							
Hospitalizations							
Home care							
Direct non-medical costs							
Transportation							

Paid caregiver time							
Indirect non-medical costs							
Patient time missed from work							
Unpaid caregiver time off from work							
CONSEQUENCES						Difference Mean (95% CI)	
Symptom impact							
Patient distress days							
Patient disability days							
Adverse Events						Difference Mean (95% CI)	
Serious adverse events							
Moderate adverse events							
Mild adverse events							

Health Related Quality of Life Impact						Difference Mean (95% CI)	
Quality Adjusted Life Years							
Quality of Life profile							

3.5.4 National Institute for Health and Care Excellence (NICE)'s medical technologies evaluation programme methods guide on cost-consequence analyses (CCAs)

In the current subsection, the guidance for cost-consequence analysis is presented, provided by the National Institute for Health & Care Excellence (NICE), the Health Technology Assessment Organization for England, which is considered the "centre of excellence" in economic evaluation techniques followed [REF-78][REF-80].

NICE mentions, which is applicable for all HTA bodies, that for a submission of evidence supporting medical technology, quantification of resources and expected outcomes associated with each treatment strategy for the relevant health care pathway will be required unless data has already been published on the subject [REF-79].

A CCA assesses costs and resource consequences and clinical benefits resulting from, or associated with, i.e., as a consequence of the use of the intervention and comparator under evaluation.

The range of costs and resource consequences to be included in the analysis depends on the clinical characteristics of individual medical technologies and the comparator. Generally, the following apply:

- Typically, cost-consequence analyses include calculating and presenting estimates of resource use and of clinical benefits as separate domains of the evaluation.
- Estimates of resource use should include comparative costs of technology (and infrastructure) acquisition, use and maintenance. Focusing on these costs may be particularly applicable when the clinical effects of the intervention can be assumed to be almost the same as those of comparator technologies.
- Resource use estimates may be informed by health care service use outcome quantification, such as length of hospital stay, or number of hospitalisations, outpatient or primary care consultations associated with the use of the technology or its comparators.

The aim of the study determines the construction and assumption of any analysis. Models describing the disease aetiology/care pathway route as discrete states, through which a patient will transition, to allow costs and effects to be appropriately assigned, quantifying the operational costs and effect of introducing an intervention. Model time horizons for accrual of effects and costs should be stated [REF-78][REF-79][REF-81].

A discount rate of 3.5% is typically applied in the UK. Other countries may have different discounting rates.

Model perspective should state whether funder or social perspective is being adopted. The default analysis will be a funder perspective. Amortisation of one-off costs will be assumed over 8 years, unless otherwise indicated, to allow estimation of cost per annum which should be divided the expected number of patients visits to estimate a pro-rata charge per use. Amortisations periods may vary between locations.

If an intervention is used on multiple conditions, the sponsor/pilot should present each condition's costs and utility as separate analysis.

Uncertainty analysis techniques (relating to chance, evidential and model uncertainty) such as scenario-based deterministic sensitivity analyses, threshold analyses or probabilistic sensitivity analyses can be added to a CCA. The level of complexity should be appropriate for the development stage of a technology, the specific technology and its comparator health care pathway.

Where strategies have near equivalence in performance e.g., clinical economic or health economic, analysis may focus on the non-equivalent performance domains.

4 Objectives & Key Performance Indicators of HosmartAI pilots

4.1 Linking Pilot's Objectives with KPIs

In order to be able to identify the KPI of each technology, the first step is the objective setting. In the tables below (from Table 9 to Table 19) the primary and secondary objectives of each pilot are presented linked with the associated KPI categories from the 5 HosmartAI KPI pillars (clinical, PROMs, PREMs, Productivity/Efficiency, Economic). At the time of preparation of the current report, the actual PROM and PREM metric selection on behalf of the pilots had not been finalized yet, hence only the KPI categories of each pilot have been recorded. The specific KPI metrics are going to be presented at Deliverable 5.3. Each pilot will be able to report the specific questionnaire/instrument of choice upon protocol approval from the hospital's scientific committee.

The recommendations from the health economic experts are the following:

- Clinical Outcome (at least 1 endpoint)
- PROM (EQ-5D-3L and/or a disease-specific questionnaire depending on the disease area in scope)
- PREM (either disease/domain-specific or LIKERT scale or Visual Analogue Scale)
- Productivity measures (average length of stay, diagnostic accuracy, average waiting time, bed occupancy rate, avoidable admissions). Productivity measures are heavily dependent on the kind/type of each technology.
- Economic (direct medical / non-medical costs and indirect costs).

Regarding the analysis, the recommendation for all pilots is to be performed incrementally, as discussed in Chapters 2 and 3 for both outcomes and costs.

It should be noted that most of the pilots do not report in their objectives the Economic KPIs since it is not necessary for the hospital scientific committee approval. Still, the economic aspects of all technologies will be gathered to perform the cost consequence analysis as reported in Chapter 3 and DoA.

4.2 Pilot 1 Objectives & Respective KPIs

Pilot #1: Development of a clinician-friendly, interpretable computer-aided diagnosis system (ICADx) to support and optimise clinical decision making in multi-specialty health care environment

4.2.1 Pilot 1 Medical Scenario #1

Echocardiography for assessment of cardiac function

The number of patients receiving their initial care and diagnostic management in emergency departments (ED) is increasing. The standard diagnostic process for patients with suspected cardiovascular diseases in the ED includes patient's current history, focusing on their most acute symptoms that led them to the emergency room, and previous history, focusing on disease states that may be related to their current condition, clinical examination, electrocardiography, chest X-ray and laboratory exams (including high-sensitive troponin assays). Based on the initial findings the physician decides if the patients should be hospitalized, as well as if an acute management should be decided, i.e., in acute coronary syndromes. All patients undergo a full echocardiographic study. Echocardiography, i.e., an ultrasound of the heart, is a widespread imaging technique that is routinely used to assess cardiac morphology and function. In particular, the assessment of Left Ventricular (LV) systolic function is important for diagnosis, management, follow-up, and prognostic evaluation of patients with heart problems. LV assessment involves the measurement of the Left Ventricular Ejection Fraction (LV-EF), which is the ratio of change in the left ventricular end-systolic and end-diastolic volumes. End-diastole and end-systole refer to the beginning and ending of the cardiac contraction, respectively. Accurate measurement of the LV-EF, also termed as EF, is critical, as it has been shown to be highly correlated with morbidity and mortality [REF-82][REF-82]. Conventionally, LV-EF measurement requires manual tracing (classic "biplane" Simpson's approach) of the left ventricle in echocardiogram frames at the end-systole and end-diastole phases. Despite being performed by experienced cardiologists, the interpretation of echocardiogram images is a highly time-consuming process and suffers from high variance [REF-83], attributable to human subjectivity and irregularity of the heart cycles. Finally, echocardiography, clinical and laboratory results are analysed and used in order to design the therapeutic management. In Table 9 the KPI categories of the first medical application scenario are presented.

Table 9: Pilot #1. First Medical Application Scenario.

Objectives	**KPI category / Metric**	**Respondent**
The development & evaluation of an AI-driven, clinician-interpretable computer-aided diagnostics/detection suite (ICADx/ICADe) to support the cardiologist' decision making related to timely diagnosis, early symptom screening, patient risk stratification and		

Objectives	KPI category / Metric	Respondent
guidance to optimized prevention or treatment options. **Pilot 1.1** **ECHO cardio fractions – abnormality id & diagnosis**		
Primary Objectives		
To determine if the automatic estimation of left ventricular (LV) ejection fraction (EF) and global longitudinal strain (GLS) from echocardiography (ECHO) scans by the artificial intelligence (AI)-based tool is non-inferior in terms of accuracy to semi-manual estimations by cardiologists of various experience levels.	Clinical effectiveness	Clinician
To determine if the combination of the cardiologist and the AI-based tool is superior to the routine clinical practice of the cardiologist in terms of correct diagnosis of left ventricle function (normal or abnormal) based on examination of ECHO scans, accounting for the cardiologist's level of experience.	Clinical effectiveness	Clinician
Secondary Objectives		
To evaluate the time required for LV-EF and LV-GLS measurement from ECHO scans by the AI-based tool and by semi-manual estimations of cardiologists of various experience levels.	PRODUCTIVITY	Clinician
To evaluate the time required for diagnosis with and without the assistance of the AI-based tool.	PRODUCTIVITY	Clinician

Objectives	KPI category / Metric	Respondent
To evaluate the usability of the AI-based tool based on cardiologists' feedback after use (via System Usability Scale questionnaire).	PREM/UREM	Clinician

4.2.2 Pilot 1 Medical Scenario #2

Capsule Endoscopy for Small Bowel Disorders

In patients under consideration for Capsule Endoscopy (CE), initial assessment typically includes symptom evaluation, laboratory assessment, and endoscopic procedures, as well as cross-sectional imaging (e.g., magnetic resonance enterography) in selected patients. For patients who have documented overt gastrointestinal (GI) bleeding (excluding hematemesis) and negative findings on high-quality EsophagoGastroDuodenoscopy (EGD) and colonoscopy, CE is performed as the next diagnostic step. CE can show additional findings in patients with prior negative endoscopic and imaging studies. In retrospective and prospective case series, the diagnostic yield of CE was 50%–72% in patients with obscure overt bleeding. In a retrospective cost-effectiveness study, the use of CE in patients with obscure bleeding had a higher diagnostic yield than other imaging procedures, and was associated with a lower cost per positive diagnosis. Based on the evidence of a relatively high diagnostic yield with CE, the American Gastroenterological Association (AGA) consensus recommended CE to be performed, rather than radiographic studies or angiography, in hemodynamically stable patients with overt bleeding. In those patients who are hemodynamically unstable, more urgent radiologic studies (e.g., angiography) may be more appropriate than CE. In patients with an overt, obscure bleeding episode, it is recommended CE to be performed as soon as possible. Because diagnostic yield appears to decrease with each day of delay, but optimal timing has not been defined definitively, CE is recommended to be performed as soon as possible within the first 24 hours in patients with ongoing overt bleeding after prior emergency negative studies. Finally, although CE is an effective non-invasive method to examine small intestine disorders, it suffers long review times [REF-84] for the busy GI department, it might lead to missed suspicious lesions and there is a need for experienced physicians to interpret its findings [REF-85]. In Table 10 the KPI categories of the second medical application scenario are presented.

Table 10: Pilot #1 Second Medical Application Scenario.

Objectives	**KPI category / Metric**	**Respondent**
The development & evaluation of an AI-driven, clinically interpretable computer-aided diagnostics/detection suite (ICADx/ICADe) to support the gastroenterologist's decision making related to timely diagnosis, early symptom screening, patient risk stratification and guidance to optimized prevention or treatment options **Pilot 1.2** **Capsule endoscopy small bowel disorders – abnormality identification and diagnosis**		
Primary Objectives		
To determine if the combination of the gastroenterologist and the AI-based tool is superior to the gastroenterologist alone in terms of primary diagnosis of small bowel conditions via VCE, accounting for the gastroenterologist's level of experience.	Clinical effectiveness	Clinician
Secondary Objectives		
To evaluate the accuracy of the AI-based tool in detecting small bowel abnormalities in CE videos.	Clinical effectiveness	Clinician
To evaluate the accuracy of the AI-based tool in classifying small bowel abnormalities identified in CE videos.	Clinical effectiveness	Clinician
To determine if the combination of the gastroenterologist and the AI-based tool is superior to the gastroenterologist alone in detecting small bowel abnormalities in CE videos, accounting for the gastroenterologist's level of experience.	Clinical effectiveness	Clinician

Objectives	KPI category / Metric	Respondent
To determine if the combination of the gastroenterologist and the AI-based tool is superior to the gastroenterologist alone in classifying small bowel abnormalities in CE videos, accounting for the gastroenterologist's level of experience.	Clinical effectiveness	Clinician
To determine if the combination of the gastroenterologist and the AI-based tool is superior to the gastroenterologist alone in terms of time required for examining CE videos, accounting for the gastroenterologist's level of experience.	Productivity	Clinician
To evaluate the usability of the AI-based tool based on gastroenterologists' feedback after use (via System Usability Scale questionnaire).	PREM/UREM	Clinician

4.2.3 Pilot 1 Medical Scenario #3

Coronary Computed Tomography Angiography for Coronary Artery Disease

A conventional routine in clinical practice over the years has been to employ validated diagnostic models of the pre-test probability (PTP) of stable, albeit obstructive, Coronary Artery Disease in order to direct downstream testing. After the first screening, adult patients with low to intermediate PTP undergoing coronary computed tomography angiography (CCTA), including calcium scoring, because of suspected Coronary Artery Disease. Most existent models have modest performance (with remarkable overestimation of risk in certain subgroups such as women) while very few studies have data regarding the effect of PTP-based models on clinical decision-making regarding further testing or patient outcomes. Practice guidelines for the management of stable chest pain from the European Society of Cardiology (ESC) are congruent in their recommendations for the use of CCTA as a first-line diagnostic option in symptomatic individuals deemed to be at a low to intermediate pre-test likelihood of having obstructive Coronary Artery Disease. However, in day-to-day clinical practice, a significant number of individuals undergoing CCTA have minimal or no Coronary Artery Disease. As a direct consequence of the expanding use of CCTA, there is a growing interest within the medical community regarding ways to optimize patient selection with the goal of

improving diagnostic yield of CCTA utilization within the context of clinical practice. Hence, there is a need for clinically based models that can predict the PTP of stable Coronary Artery Disease and as a result function as gatekeepers to identify low-risk individuals who are unlikely to have obstructive Coronary Artery Disease and unlikely to need further diagnostic testing. In Table 11 the KPI categories of the third medical application scenario are presented.

Table 11: Pilot #1 Third Medical Application Scenario.

Objectives	**KPI category / Metric**	**Respondent**
The development & evaluation of an AI-driven, clinician-interpretable computer-aided diagnostics/detection suite (ICADx/ICADe) to support the cardiologist' decision making related to timely diagnosis, early symptom screening, patient risk stratification and guidance to optimized prevention or treatment options. **Pilot 1.3 – CAD abnormality identification and diagnosis**		
Primary Objectives		
To evaluate AI-based tool's ability to provide an accurate assessment of the presence or absence of obstructive coronary artery disease.	Clinical effectiveness	Clinician
Secondary Objectives		
To evaluate clinicians' satisfaction regarding the AI-based tool, based on the feedback after use.	PREM	Clinician
To evaluate if there will be a reduction in diagnostic time for a clinician with or without using the AI-based tool.	Productivity	Clinician

4.2.4 Pilot 1 Medical Scenario #4

Pregnancy Abnormality Detection

The most common indications for referrals and admissions of pregnant women in a tertiary obstetric clinic with a High-Risk Pregnancy Unit (HRPU) are threatened preterm labour, ischemic placental disease (mostly hypertensive disorders and fetal growth restriction) and finally, poorly controlled hyperglycaemia in pregnancy. These conditions are increasing in incidence in high-resource countries as advanced maternal age and obesity become more common. The patients will either attend the primary or secondary settings on a regular appointment where the health professionals will assess their exams or come to the emergency unit with a symptom. In the first scenario, health professionals need to decide on examinations already performed, most commonly ultrasound of the fetal or a recording of blood pressure or blood sugar levels that categorize the pregnancy as high-risk. They are faced with a dilemma of urgently referring the patient to a tertiary centre or on a scheduled appointment basis. Often, they resort to the first choice and this is usually unnecessary. In the second scenario, a pregnant woman will present with symptoms of threatened preterm labour, vaginal bleeding or reduced fetal movements, again creating uncertainty on a possible adverse outcome. Health professionals often assess correctly the patient but still refer to a tertiary centre due to their own anxiety. This situation is associated with increased rates of anxiety of the pregnant women and their environment, leads to unnecessary examinations and admissions and therefore consumes valuable resources. A cost-effective management of these conditions in pregnancy may be achieved by targeted training of health professionals, appropriate available equipment in primary and secondary settings and most importantly a clear pathway of communication with a multidisciplinary maternal-fetal medicine team in the tertiary centre, allowing the triage of patients at the primary site. The current COVID-19 epidemic has clearly shown that this approach may prove even more crucial in times of crisis. This approach is targeted to improve efficacy and reduce the anxiety of health professionals, along with reducing the stress of pregnant women and unnecessary use of resources. More importantly, it is also expected that it will improve pregnancy outcomes. In Table 12 the KPI categories of the fourth medical application scenario are presented.

Table 12: Pilot #1 Fourth Medical Application Scenario.

Objectives	**KPI category / Metric**	**Respondent**
The development & evaluation of an AI-driven, clinician-interpretable computer-aided diagnostics/detection suite (ICADx/ICADe) to support the clinical decision making related to timely diagnosis, **early symptom screening, patient risk stratification and guidance to optimized prevention or treatment options.**		

Objectives	KPI category / Metric	Respondent
Pilot 1.4 – pregnancy abnormality detection		
Primary Objectives		
To evaluate the performance of the AI-based tool to be developed in terms of identification of preterm labour and/or fetal growth restriction cases	Clinical effectiveness	Clinician
Secondary Objectives		
To evaluate clinicians' satisfaction regarding the AI-based tool, based on the feedback after use.	PREM	Clinician
To evaluate whether diagnostic time for the pregnancy abnormality diagnostic process is less (duration) with the AI-based diagnostic tool to diagnosis without the tool.	Productivity	Hospital administrative staff
To evaluate if use of AI-based tools reduces unnecessary use of resources.	Productivity	Hospital administrative staff / Clinician

4.3 Pilot 2 Objectives & Respective KPIs

Pilot #2: Optimizing the use of radiotherapy

The Radiotherapy Unit at CHU de Liège face organisational challenges in dealing with a considerable number of patients. Type, aggressivity, size and location of the tumour are among the relevant parameters for an appropriate radiotherapy. Although the specific characteristics of the tumour and the patient's physical conditions are the roots of all radiotherapy, caregivers cannot ignore oncological treatments and psychosocial context. Thus, before starting the radiotherapy, an accurate patient-centred planning considering all these parameters is needed, while following guidelines to ensure the right treatment at the right time on the right treatment machine throughout the entire course of the treatment. The availability of the machines and of human resources, not to mention patient prefe rences and travel possibilities are among the obstacles to optimize scheduling. To remedy this problem,

is necessary to accurately weigh each one of dozens of variables according to the oncological, psychosocial and organisational context of the patients, in order to offer them the best possible treatment schedule with a solution reaching Pareto optimality [REF-86].

The goal is to establish an AI algorithm for optimizing patient scheduling [REF-87][REF-88] in a context of patient-centred planning where variable weighting and resource availability may change. It is expected that the AI system will take into account i) all variables as electronically present in the patient record; ii) electronic forms completed by the consulting radiation oncologists; iii) treatment machine characteristics, iv) scheduling for other patients as present in the patient management system used by the radiotherapy department (Mosaiq) [REF-89], v) patient preferences thanks to a chatbot. In Table 13 the KPI categories of pilot 2 are presented.

Table 13: Pilot #2 Hospital Patient Scheduling with the Use of AI.

Objectives	**KPI category / Metric**	**Respondent**
AI management of multiple scheduling pressures e.g., up to 10 factors considered to find growing numbers of patients an app. Human processing ability has limits. This pilot's objective is to establish an AI algorithm, housed in a robot and in an app, for optimizing patient scheduling and outcomes in the context of patient-centred planning. **Pilot 2 AI-based scheduling**		
Primary Objectives		
To evaluate whether scheduling managers and clinicians' satisfaction with the process of scheduling radiotherapy appointments is superior with the AI scheduling SW with and without Chatbot, then through a human scheduling manager.	PREM	Clinician/ scheduling manager
To evaluate whether patients' satisfaction with the process of scheduling radiotherapy appointments is superior with the AI scheduling SW than	PREM	Patient

Objectives	KPI category / Metric	Respondent
through a scheduling manager, with and without Chatbot.		
Secondary Objectives		
To determine if the combination of the use of AI scheduling SW has lower overall treatment cost compared to a scheduling manager, by alignment of resources (staff, radiotherapy suite, other) to patient appointments – revised and otherwise.	Economic	Hospital administrative staff

4.4 Pilot 3 Objectives & Respective KPIs

Pilot #3: Treatment Improvement with the use of innovative technologies and robotics in rehabilitation process

At IRCCS San Camillo rehabilitation therapies are delivered both as conventional care by a dedicated Clinical Service for Neuromotor Rehabilitation and as experimental modalities by a dedicated research laboratory (Laboratory of Rehabilitation Technologies). Innovative modalities include a variety of technology-based approaches (e.g., virtual reality, inertial measurement units, robotics).

So far, almost 90% of the rehabilitation care is provided in one-to-one (patient/therapist) settings and scheduled on weekly basis, by collecting information from each single Operative Units of the Neurorehabilitation Department and manual allocation of treatments according to resources available. Moreover, eventual adverse events and the delivery of services can be registered only manually, being dependent on the presence of the therapist.

As a state-of-art, each technology operates like an independent environment where all data is collected and stored, with possibility to have access to raw data and/or to synchronise them with the clinical history of each patient. Currently, the best standard commercially available allows just backup data from different devices of the same company, in a common repository (physical or cloud) to avoid information lost, in case of hard storage corruption. In Table 14 the KPI categories of pilot 3 are presented.

Table 14: Pilot #3 Rehabilitation of Patients with Neurological Problems.

Objectives	**KPI category / Metric**	**Respondent**
Environment and behaviour model imprinting of robotic rehabilitation aids for the rehabilitation of neurological patients from hospital to home. **SW+HW mediated rehabilitation of patients with neurological problems** **Pilot 3**		
Primary Objectives		
To determine whether a behaviour & environment imprinted robotic aid is superior to a physiotherapist in terms of rehabilitation of patients with neurological conditions across the measures listed here: - Trunk Control Test (TCT) - Reaching Performance Scale (RPS) - Box and Blocks Test (BBT) - Nine-hole Pegboard Test (NHPT) - Berg Balance Scale (BBS) - 10 meters walking test - Functional Ambulation Categories (FAC)	Clinical Effectiveness	Clinician
Secondary Objectives		
To determine whether the robotic aid had a lower cost of rehabilitation than a physiotherapist.	Economic	Hospital administrative staff
To evaluate whether the number of patients successfully rehabilitated is higher with the robotic aid than through a physiotherapist without the robotic aid.	Productivity	Hospital administrative staff

4.5 Pilot 4 Objectives & Respective KPIs

Pilot #4: Robotic Systems for minimally Invasive Operation

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia in adults and one of the 3 cardiovascular pandemics of the XXI century according to the World Health Organization (WHO). The currently estimated prevalence of AF in adults is between 2% and 4%, and a 2.3-fold rise is expected, owing to extended longevity in the general population and intensifying search for undiagnosed AF. This disease is associated with substantial morbidity, (5-fold increased risk of stroke and 3-fold increased risk of heart failure) and mortality (2-fold risk of death). In addition, it is also associated with significant health care cost burden mostly associated with hospitalizations. Most treatments are aimed to reduce the complications of this disease, like anticoagulation to prevent stroke. However, pharmacological therapies to prevent AF episodes or the progression of the disease have shown disappointing results with minor or neutral effects in major outcomes. Catheter ablation has shown more effective than drugs to keep patients without this arrhythmia, to prevent its progression and to reduce follow-up events. However, ablation is an invasive treatment which is associated with a small but still significant risk of complications and with suboptimal results because the precise AF mechanism and best mapping approach are still poorly defined. In addition, it is mostly performed by manual operation of catheters which require a significant amount of dexterity and experience. This results in substantial heterogeneity of clinical practice and important barrier to offer this therapy to many patients. As an example, 5164 AF ablation procedures were performed in Spain in the year 2019 which means that only 0.1% of Spanish patients with AF had the opportunity to receive this treatment that year. Robotic systems aimed to reduce the learning curve required to perform AF ablation procedures, to reduce complications and to make the procedure less operator dependent and more automatized have been developed in the past. However, most of these systems were based on mechanical navigation of the catheter with little impact on safety and operator dependency or were based on permanent magnets which resulted in slow operation, long procedures and difficulties to give the catheter enough contact force with the tissue. In Table 15 the KPI categories of pilot 4 are presented.

Table 15: Pilot #4 Remote vs. AI Navigation in Catheter Tip Ablation.

Objectives	**KPI category / Metric**	**Respondent**
AI in remote navigation – a comparison of semi- and fully- AI-, versus manually-, guided catheter tip ablation of arrythmia triggers. The intervention will include a comparison of manual and AI mapping,		

Objectives	KPI category / Metric	Respondent
remote and present undertaking of tasks for both in-vitro automated and reference manual approaches. **Pilot 4**		
Primary Objectives		
To evaluate the **User's satisfaction with robotic approach.**	UREM (**SUS - System Usability Scale)**	Clinician
Secondary Objectives		
To determine whether the **accuracy of automatic and semi-automatic AI guided-ablation is inferior to manually guided-ablation for atrial fibrillation patients** undergoing surgery	Clinical effectiveness	Clinician
To determine whether the accuracy and duration of **remote fully- and semi-automatic AI guided-ablation is worse than manually guided- ablation for atrial fibrillation patients** undergoing surgery	Clinical effectiveness	Clinician
To determine whether the **time to ablate multiple arrythmia targets is >= 25% shorter** with the AI fully- and semi-automated approach than with the manual approach.	Productivity	Hospital administrative staff

4.6 Pilot 5 Objectives & Respective KPIs

Pilot #5: Assistive Care in Hospital: Robotic Nurse

UM deals with exploiting AI's and technologies to increase patient satisfaction and optimization of hospital resource utilization by delivering a more personalized care. Introducing robotics in health care can compensate for the shortage in the human workforce. Robots and ICT can implement passive sensing (e.g., telemonitoring, recognition of moods and (psychological) symptoms), and virtual support (e.g., companion mode and empathy) and offer i) more time for professionals to provide care rather than miscellaneous and repetitive

tasks or administration ii) new types of data and new means of representation to be included in the decision making during the regular clinical workflow. The overall objective of the pilot is to develop a social robotic system (SRS) that **supports nursing and care** through automated data collection, **improves decision making during clinical workflow** by aggregation and efficient representation of relevant patient data during regular grand rounds (doctor visits) and **improves the quality of care and patient experience** via companion functionality. In Table 16 the KPI categories of pilot 5 are presented.

Table 16: Pilot #5 Robotic Nurse.

Objectives	**KPI category / Metric**	**Respondent**
Assistive robot care on Grand Rounds and Thoracic surgery patient assistance through robot nurse in the hospital **Pilot 5**		
Primary Objectives		
To determine whether patients' objective vital signs, - blood pressure, temperature, hydration, are recorded, were poorer with the robot nurse than without the robot nurse.	Clinical-effectiveness	Hospital staff
Secondary Objectives		
To determine whether patients' subjective measures - emotions, pain, PROMs, are improved with the robot nurse than without the robot nurse.	PROM	Patient
To determine whether the use of robots within grand round improves the time of preparation (less duration).	Economic	Hospital administrative staff
To determine whether the overall cost of patient care was higher with the virtual assistant than without it.	Economic	Hospital administrative staff
To evaluate whether the self-efficacy and workload of staff is higher with robot nurse than with the robot nurse.	Productivity	Hospital administrative staff
To evaluate whether the interaction time, by non-urgent and other types of	Productivity	Nurse

Objectives	KPI category / Metric	Respondent
interaction, is greater with the robot nurse than without the robot nurse.		
To determine whether the user experience (UEQ), patient engagement (PHE) and perceived quality of medical care (PQMC) is greater with the robot nurse, than without the robot nurse.	PREM	Patient
To determine whether the usability of the CDSS system is superior with the robot nurse than without the robot nurse (System Usability Scale)	UREM	Nurse
To evaluate whether the treatment of patients with the CDSS system is less acceptable than without the CDSS system.	PREM	Patients
To determine whether the quality-adjusted life-year is greater with than without the robot nurse. (EQ-5D-3L)	PROM	Patients

4.7 Pilot 6 Objectives & Respective KPIs

Pilot #6: Assistive Care in Care Centre: Virtual Assistant

INTRAS memory clinics and neuro-psychological rehabilitation centres work both in the outpatient setting and at home, with a catalogue of services involving Neuropsychological assessment, Cognitive stimulation and rehabilitation, Active aging programs, Psychomotricity program, Speech therapy, Multisensory Stimulation Therapy, Training in instrumental activities of daily living, and Psychotherapeutic and psychoeducational programs for the caregiver. These care services aim for cognitive, physical, social and emotional health and wellbeing, making already use of digital tools (e.g., remote sessions, Suite Gradior, VR) as an important resource supporting the therapeutic plans.

Despite the innovative culture in prevention, rehabilitation and assistive care for older adults there is still a gap in systematizing information collection processes from the patient journey to allow greater degree of personalization. Each technology usually operates as an independent environment and is not significantly endowed with aspects of personalization and decision support. In addition, the low communication among public-private sectors and

between social-health services sometimes do not facilitate integrated case management and care coordination, making difficult early detection, adequate holistic monitoring progress or high personalization of health and social prescriptions.

Alternative solutions are required to also address the concerns of citizens for ageing-in-place, with a growing demand for innovative services aimed to provide care that is both affordable and meets the emotional, social, cognitive and physical needs of older adults, while can also be an answer to the increasing ratio of older adults living alone or suffering unwanted loneliness, one of the epidemics of the century, implying risk of premature death, worsening of health, physical, cognitive deterioration and loss of quality of life of the elderly. At the same time, it is important to attend to the difficulty balancing sustainability and resources with high demand and the need for adequate intensity and quality of services. In Table 17 the KPI categories of pilot 6 are presented.

Table 17: Pilot #6 Virtual Assistance in Rehabilitation Centres.

Objectives	**KPI category / Metric**	**Respondent**
A virtual assistant to screen and apply personalized treatment modalities according to cognitive need to moderate cognitive decline as well as detect the presence of frailty, in care centres and private homes. **Pilot 6**		
Primary Objectives		
To determine whether the virtual assistant has a greater effect than current care practice in preventing cognitive decline in elderly adults, in care centres and home settings, using *Mini Mental test	Clinical effectiveness	Patient
To determine whether the virtual assistant has a greater effect than care practice in preventing falls in patient mood, in care centres and home settings using *Geriatric depression scale, GDS of Yesavage	Clinical effectiveness	Patient
To determine whether detection of patient frailty (present/absent) is	Clinical effectiveness	Patient

Objectives	KPI category / Metric	Respondent
superior with than without the virtual assistant, in care centres and home settings.		
To determine whether the patient remains at home for less time, according to their degree of frailty, with or without the virtual assistant.	Clinical effectiveness	Patient
Secondary Objectives		
To determine whether the overall cost of patient care was lower with the virtual assistant than without it.	Economic	Care Center administrative staff
To evaluate whether the staff/patient ratio was higher with the virtual assistant than without it.	Productivity	Care Center administrative staff
To evaluate patients' quality of life (QALYs) with and without the virtual assistant, over the duration of the trial (EQ5D-3L).	PROM	Patient
To determine whether patients found the care experience more acceptable, easier to work with and more useful, with or without the virtual assistant.	PREM	Patient

4.8 Pilot 7 Objectives & Respective KPIs

Pilot #7: Smart Cathlab Assistant

Percutaneous Coronary Intervention (PCI) is an image guided procedure used to treat a narrowing of the coronary arteries of the heart by placing a stent to widen the blood vessel diameter. Currently, X-ray imaging is typically used during such procedures for navigation, but other data sources like ultrasound imaging and blood flow measurements are often included, in particular to treat complex cases.

The integration of multiple imaging and data sources leads to data clutter and makes it difficult for a clinician to interpret the data and extract meaningful insights to diagnose and

treat the patient. Image interpretation is usually done manually, requiring highly skilled experts and may lead to fatigue and errors. To assist the clinician in the understanding and assessment of clinical data there is a need for smart clinical applications that are able to automatically interpret medical images and do a quantitative assessment where possible. Computer-assisted interpretation of images could go a long way in further improving accuracy, offering a helping hand to the clinician. In Table 18 the KPI categories of pilot 7 are presented.

Table 18: Pilot # 7 Coronary Angiogram.

Objectives	**KPI category / Metric**	**Respondent**
To pilot an AI built registry of coronary angiogram & physiologic evaluations to provide imagery and guidance to Cathlab operations. **Pilot 7**		
Primary Objectives		
To determine whether the AI automated coronary angiogram image interpretation has inferior accuracy to image interpretation by clinicians without the AI SW.	Clinical effectiveness	Clinician
To determine whether a greater incremental difference between post- and pre- coronary operation resting index measure (iFR/RFR) or hyperemic index (FFR) is observed using AI-facilitated imagery labelling and operational guidance than with manual imagery labelling and guidance, by manual or a motorized wire.	Clinical effectiveness	Clinician
To evaluate intravascular imaging data of patient evaluated by either Intravascular ultrasound (IVUS) Optical coherence tomography (OCT) technique before and after a coronary intervention with and without AI guided coronary surgery.	Clinical effectiveness	Clinician

Objectives	KPI category / Metric	Respondent
To evaluate whether the accuracy of AI facilitated classification of patient imagery is greater than the manually facilitated classification of Coronary CT data, including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention (GE revolution and Heartflow software)	Clinical effectiveness	Clinician
Secondary Objectives		
None specified		

4.9 Pilot 8 Objectives & Respective KPIs

Pilot #8: Prognosis of cancer patients and their response to treatment combining multi-omics data

VUB focuses on accurate glioma diagnosis, which contains two major elements: (1) segmentation and (2) characterisation of the tumour, including both the determination of the subtype and grade. In this pilot, we address the ability to connect researchers with clinicians in a 'rapid learning health care' approach. To do so, requires a digital health research platform, where multimodal data and advanced analytics are integrated for the analysis of brain tumours, and a decision support system based on integrated molecular and image level research on the tumour that directly connects to, and informs, clinicians.

The unique expertise and data present at the VUB and the UZ Brussel is here leveraged to create a general framework to store and analyse raw medical data, both at the image and molecular level, in relation to brain tumours, their clinical behaviour and response to therapies. The platform offers an integrated view on the patient data for research, while conforming to GDPR and patient legislation, thus enabling AI-driven extraction of new information on such tumours. An important feature of the platform is the tight integration with the hospital information system. Extraction of data from the latter is performed (semi-) automatically, allowing to iteratively update the data collection, retrain models and deploy them in the clinical setting. In Table 19 the KPI categories of pilot 8 are presented.

Table 19: Pilot # 8 AI Management of Glioma Patients.

Objectives	**KPI category / Metric**	**Respondent**
To create a digital research platform for Glioma management by combining genetic and image data. The platform offers an integrated view on the patient data for research, while conforming to GDPR and patient legislation, thus enabling AI driven extraction of new information on such tumours. A surgical and data collection element is included where extensive patient data is collected to achieve the above aim. The AI mediated tools will not be directly measurable in terms of their accuracy in relation to manual procedures but are instead intended to highlight aspects of the data that require particular attention during the oncology consult. **Pilot 8**		
Primary Objectives		
To provide integrated view on Glioma patient data (genetic & imaging) of tumour detection leading to better diagnosis in comparison to current technology.	Clinical effectiveness	Clinicians and Researchers
To investigate the effectiveness of integrated data analysis (genetic and images) for better tumour detection, leading to improved treatment and better survival versus the current clinical practice. (Improvement in Overall Survival (OS) and/or Progression Free Survival (PFS))	Clinical effectiveness	Clinicians
Secondary Objectives		
To investigate the usability of the AI technology vs. the current technology (System Usability Scale)	UREM	Clinicians
To investigate the duration of the new technology versus the current one	Productivity	Clinicians

5 Conclusions

Effective Artificial Intelligence algorithms and robots for health care in hospitals, care homes and people's homes presents potential contributions to managing predicted demographic trends of a decrease in EU working age population and a commensurate expansion in the >80 years olds [REF-90]. With such a change and associated frailty or older people, health care needs are likely to expand e.g., for mobility, care, rehabilitation and other needs. The use of safe interventions to resolve these challenges will be a major familiarisation change for individuals and society. For example, hospital managers considering investing in net enabled AI services, may need to maintain sufficiently trained staff in case eventualities e.g., logistic supply problems, network availability etc. COVID has taught us the value of planning for resilience. And yet, the intervention will act as a layer of resilience e.g. in case of staff sickness, potential lowering patient waiting times, adding a mutually reinforcing contributor to health solutions.

This report has described the framework of performance management, the background of KPI need and types and undertaken an analysis of HosmartAI pilot objectives, intended KPIs and measures, illustrating the link identified need and measurement of solution for extant needs/opportunities.

The KPIs analysed cover the whole spectrum of the health care sector, both final and intermediate clinical endpoints, patient reported outcomes and experience measures, economic and productivity endpoints. The objective was to be able to cover all involved stakeholders of the health care sector with the KPIs, namely, medical practitioners, patients, hospital managers and health care policy makers.

We expect from the KPI set and tests described, a set of outcomes which should demonstrate meaningful value to hospital staff, patients and industry, not least from the embedded use of KPIs to measure outcomes of projects. Hopefully, the current report will shed light on the economic evaluation of the Artificial Intelligence field and could be used as a point of reference for future analysis in the respective field.

6 References

[REF-01]	Dilsizian SE, Siegel EL. (2014) Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 16(1):441. doi: 10.1007/s11886-013-0441-8. PMID: 24338557.
[REF-02]	Jha S, Topol EJ. (2016) Adapting to Artificial Intelligence: radiologists and pathologists as information specialists. JAMA 13;316(22):2353-2354. doi: 10.1001/jama.2016.17438. PMID: 278989316:2353–4.
[REF-03]	Lee CS, Nagy PG, Weaver SJ, et al. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol 201(3):611-7. doi: 10.2214/AJR.12.10375. PMID: 23971454.
[REF-04]	Neill, D.B. (2013). Using Artificial Intelligence to Improve Hospital Inpatient Care. IEEE Intelligent Systems, 28, 92-95. doi:10.1109/MIS.2013.51
[REF-05]	Cole R.E., Scott W.R. (2000) The quality movement & organization theory. Sage Publications.
[REF-06]	Institute of Medicine (US). (2001) Committee on Quality of Health Care in America, Crossing the quality chasm: A new health system for the 21st century. Washington (DC): National Academies Press (US) DOI: 10.17226/10027 PMID: 25057539.
[REF-07]	Grol R. (2001) Improving the quality of medical care: Building bridges among professional pride, payer profit, and patient satisfaction. JAMA 286, 28;286(20):2578-85. doi: 10.1001/jama.286.20.2578. PMID: 11722272.
[REF-08]	Raven J.H., Tolhurst R.J., Tang S., van den Broek N., (2012) What is quality in maternal and neonatal health care? Midwifery 28(5):e676-83. doi: 10.1016/j.midw.2011.09.003. Epub 2011 Oct 20. PMID: 22018395.
[REF-09]	Institute of Medicine (US) Committee to Design a Strategy for Quality Review and Assurance in Medicare. (1990) Medicare: A Strategy for Quality Assurance: Volume 1. Lohr KN, editor. Washington (DC): National Academies Press (US); DOI: 10.17226/1547 PMID: 25144047.
[REF-10]	Nordgren L. (2009). Value creation in health care services – developing service productivity: Experiences from Sweden. Int. J. Public Sect. Manag. 22, 114–127.
[REF-11]	Longtin Y., Sax H., Leape L.L., Sheridan S.E., Donaldson L., Pittet D. (2010) Patient Participation: Current Knowledge and Applicability to Patient Safety. Mayo Clin. Proc. 85(1):53–62. doi: 10.4065/mcp.2009.0248. PMID: 20042562; PMCID: PMC2800278.
[REF-12]	EC, 2014, The 2015 ageing report — Underlying assumptions and projection methodologies, European Economy, 8/2014, European Commission, Brussels.
[REF-13]	Singh O and Kumar S. (2012) "Measuring Effectiveness of Health Program Intervention in the Field. Intelligent Information Management, 4(5):194-206. doi: 10.4236/iim.2012.45029
[REF-14]	Ramalho, A., Castro, P., Gonçalves-Pinho, M., Teixeira, J., Santos, J. V., Viana, J., Lobo, M., Santos, P., & Freitas, A. (2019). Primary health care quality indicators: An umbrella review. PloS one, 14(8), e0220888. https://doi.org/10.1371/journal.pone.0220888.

[REF-15]	Fenwick, E., Marshall, D.A., Levy, A.R. et al. (2006) Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. BMC Health Serv Res 6:52. https://doi.org/10.1186/1472-6963-6-52.
[REF-16]	Teisberg, E., Wallace, S., & O'Hara, S. (2020). Defining and Implementing Value-Based Health Care: A Strategic Framework. Academic medicine : journal of the Association of American Medical Colleges, 95(5), 682–685. https://doi.org/10.1097/ACM.0000000000000312219.
[REF-17]	Donabedian, A (2005) Evaluating the Quality of Medical Care, The Milbank Quarterly, 83(4):691-729
[REF-18]	Epstein RM., Alper BS., Quill TE. (2004) Communicating evidence for participatory decision-making. JAMA 19;291(19):2359-66. doi: 10.1001/jama.291.19.2359. PMID: 15150208.
[REF-19]	Charles C., Gafni A., Whelan T. (1997) Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). Soc. Sci. Med. 44(5):681-92. doi: 10.1016/s0277-9536(96)00221-3. PMID: 9032835.
[REF-20]	Saba G.W., Wong S.T., Schillinger D., Fernandez A., Somkin C.P., Wilson C.C., Grumbach K., (2006) Shared decision making and the experience of partnership in primary care. Ann. Fam. Med. 4(1):54-62. doi: 10.1370/afm.393. PMID: 16449397; PMCID: PMC1466999.
[REF-21]	Guadagnoli E., Ward P., (1998) Patient participation in decision-making. Soc. Sci. Med. 47(3), 329–339. doi: 10.1016/s0277-9536(98)00059-8. PMID: 9681902.
[REF-22]	Murray E., Pollack L., White M., Lo B. (2007) Clinical decision-making: Patients' preferences and experiences. Patient Educ. Couns. 65 (2), 189–196. doi: 10.1016/j.pec.2006.07.007. Epub 2006 Sep 7. PMID: 16956742.
[REF-23]	Schneider, A., Körner, T., Mehring, M., Wensing, M., Elwyn, G., Szecsenyi, J., 2006. Impact of age, health locus of control and psychological co-morbidity on patients' preferences for shared decision making in general practice. Patient Educ. Couns. 61(2), 292–298 doi: 10.1016/j.pec.2005.04.008. PMID: 15896943.
[REF-24]	Murray E., Pollack L., White M., Lo B. (2007b). Clinical decision-making: physicians' preferences and experiences. *BMC* Fam. Pract. 8, 10 https://doi.org/10.1186/1471-2296-8-10.
[REF-25]	Stevenson F.A. (2003) General practitioners' views on shared decision making: a qualitative analysis. Patient Educ. Couns. 50(3), 291–293. doi: 10.1016/s0738-3991(03)00052-1. PMID: 12900102.
[REF-26]	Towle A., Godolphin W. (1999) Framework for teaching and learning informed shared decision making. *BMJ* 319 (7212), 766–771 https://doi.org/10.1136/bmj.319.7212.766.
[REF-27]	Briggs AH, Sculpher MJ, Claxton K. (2006) Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press.
[REF-28]	Street R. L., Elwyn G., Epstein R. M. (2012). Patient preferences and healthcare outcomes: an ecological perspective. Expert Review of Pharmacoeconomics & Outcomes Research, 12(2), 167-180 doi: 10.1586/erp.12.3. PMID: 22458618.
[REF-29]	Yorkston, K, Baylor C, (2019). Tutorial - Patient-Reported Outcomes Measures: An Introduction for Clinicians, Perspectives of the ASHA Special Interest Groups, 4(1):8–15 https://pubs.asha.org/doi/abs/10.1044/2018_PERS-ST-2018-0001.
[REF-30]	Sepucha K, Ozanne E.M, (2010). How to define and measure concordance between patients' preferences and medical treatments: A systematic review of approaches

	and recommendations for standardization, *Patient Education and Counseling* 78(1):12-23 https://doi.org/10.1016/j.pec.2009.05.011.
[REF-31]	Torrance GW, Thomas WH, Sackett DL (1972) A utility maximization model for evaluation of health care programs. *Health Serv Res*. 7(2):118-33. PMID: 5044699; PMCID: PMC1067402.
[REF-32]	Blair J, McKay A, Ridyard C, Thornborough K, Bedson E, Peak M, Didi M, Annan F, Gregory JW, Hughes D, Gamble C. (2018) Continuous subcutaneous insulin infusion versus multiple daily injections in children and young people at diagnosis of type 1 diabetes: the SCIPI RCT. He*alth Technol Assess*. 22(42):1-112. doi: 10.3310/hta22420.
[REF-33]	McCabe CJ, Stevens KJ, Brazier JE. (2005) Utility scores for the Health Utilities Index Mark 2: an empirical assessment of alternative mapping functions. *Med Care*. Jun;43(6):627-35. doi: 10.1097/01.mlr.0000163666.00471.8e. PMID: 15908858.
[REF-34]	Furlong WJ, Feeny DH, Torrance GW, Barr RD. (2001) The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*. 33(5):375-84. doi: 10.3109/07853890109002092. PMID: 11491197
[REF-35]	Eadie TL, Otero D, Cox S, Johnson J, Baylor CR, Yorkston KM, Doyle PC. (2016) The relationship between communicative participation and postlaryngectomy speech outcomes. *Head Neck.* 38(1):E1955-61. doi: 10.1002/hed.24353. Epub 2015 Dec 29. PMID: 26714043; PMCID: PMC4844835.
[REF-36]	Francis, DO. (2014) Patient-centered outcomes research in voice; SIG 3 *Perspectives on Voice and Voice Disorders*, 24(1), 21–31; https://doi.org/10.1044/vvd24.1.21.
[REF-37]	Beller, JP. (2021) Early Versus Delayed Pacemaker for Heart Block After Valve Surgery: A Cost-Effectiveness Analysis, *J. Surg.Res.* 259:154-162. doi: 10.1016/j.jss.2020.11.038.
[REF-38]	Snyder A., Makarov V., Merghoub T. et. al. (2014) Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. N. Engl. J. Med 371(23):2189 https://www.nejm.org/doi/full/10.1056/NEJMoa1406498.
[REF-39]	Amtmann D, Cook KF, Johnson KL, Cella D. (2011) The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. *Arch Phys Med Rehabil.* 92(10):S12-9. doi: 10.1016/j.apmr.2011.04.025. PMID: 21958918; PMCID: PMC3668854.
[REF-40]	Cella D, Gershon R, Lai JS, Choi S. (2007) The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res.* 16(1):133-41. doi: 10.1007/s11136-007-9204-6. Epub 2007 Mar 31. PMID: 17401637.
[REF-41]	Cook KF, Monahan PO, McHorney CA. (2003) Delicate balance between theory and practice: health status assessment and item response theory. *Med Care* 41(5):571-4. doi: 10.1097/01.MLR.0000064780.30399.A4. PMID: 12719678.
[REF-42]	Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, Revicki DA, Symonds T, Parada A, Alonso J. (2008) The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res.* 17(2):179-93. doi: 10.1007/s11136-007-9295-0. Epub 2008 Jan 4. PMID: 18175207.

[REF-43]	Hawkins M, Elsworth GR, Osborne RH. (2018) Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. *Qual Life Res.* 27(7):1695-1710. doi: 10.1007/s11136-018-1815-6. PMID: 29464456; PMCID: PMC5997725.
[REF-44]	Gallop K, Nixon A, Swinburn P, Sterling KL, Naegeli AN, Silk ME. (2012) Development of a conceptual model of health-related quality of life for systemic lupus erythematosus from the patient's perspective. *Lupus.* 21(9):934-43. doi: 10.1177/0961203312441980. PMID: 22433917.
[REF-45]	Bleichrodt H, Johannesson M. (1997) Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ*. 16(2):155-75. doi: 10.1016/s0167-6296(96)00509-7. PMID: 10169092.
[REF-46]	Drummond, M. F., Stoddart, G. L., & Torrance, G. W. (1987). *Methods for the economic evaluation of health care programmes*. (First ed.) Oxford Medical Publications.
[REF-47]	Petersen MA, Groenvold M, Aaronson NK, Chie WC, Conroy T, Costantini A, Fayers P, Helbostad J, Holzner B, Kaasa S, Singer S, Velikova G, Young T; EORTC Quality of Life Group. (2010) Development of computerised adaptive testing (CAT) for the EORTC QLQ-C30 dimensions - general approach and initial results for physical functioning. Eur J Cancer 46(8):1352-8. doi: 10.1016/j.ejca.2010.02.011.
[REF-48]	Barrett, A., Hahn-Pedersen, J., Kragh, N., Evans, E., & Gnanasakthy, A. (2019). Patient-Reported Outcome Measures in Atopic Dermatitis and Chronic Hand Eczema in Adults. *The patient*, 12(5), 445–459. https://doi.org/10.1007/s40271-19-00373-y.
[REF-49]	EuroQol Group. (1990) EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 16(3):199-208. doi: 10.1016/0168-8510(90)90421-9. PMID: 10109801.
[REF-50]	Dolan P. (1997) Modeling valuations for EuroQol health states. *Med Care* 35(11):1095-108. doi: 10.1097/00005650-199711000-00002. PMID: 9366889.
[REF-51]	World Health Organization. Division of Mental Health and Prevention of Substance Abuse. (1997). WHOQOL:measuring quality of life. *World Health Organization.* https://apps.who.int/iris/handle/10665/63482.
[REF-52]	Ware JE Jr, Sherbourne CD. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30(6):473-83. PMID: 1593914.
[REF-53]	Brazier J, Roberts J, Deverill M. (2002) The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 21(2):271-92. doi: 10.1016/s0167-6296(01)00130-8. PMID: 11939242.
[REF-54]	Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. (2002) Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 15;324(7351):1417. doi: 10.1136/bmj.324.7351.1417. PMID: 12065262; PMCID: PMC115850.
[REF-55]	Cella D., Yount S, Rothrock N., Gershon R., Cook K, Reeve B, et al. (2007) The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years *Med Care*, 45(1): S3-11 http://dx.doi.org/10.1097/01.mlr.0000258615.42478.55.

[REF-56]	Aaronson NK, Acquadro C, Alonso J, Apolone G, Bucquet D, Bullinger M, Bungay K, Fukuhara S, Gandek B, Keller S, et al. (1992) International Quality of Life Assessment (IQOLA) Project. *Qual Life Res* 1(5):349-51. doi: 10.1007/BF00434949. PMID: 1299467.
[REF-57]	Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology,* 22 140, 55.
[REF-58]	Norman G. (2010) Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ Theory Pract*. 15(5):625-32. doi: 10.1007/s10459-010-9222-y. Epub 2010 Feb 10. PMID: 20146096.
[REF-59]	Sullivan GM, Artino AR Jr. (2013) Analyzing and interpreting data from likert-type scales. *J Grad Med Educ* 5(4):541-542. DOI: 10.4300/jgme-5-4-18. PMID: 24454995; PMCID: PMC3886444.
[REF-60]	Doty, M. et al. (2017), *Mirror*: International Comparison Reflects Flaws and Opportunities for Better U.S. Health Care, http://dx.doi.org/10.15868/socialsector.27698.
[REF-61]	Nuttall, D., D. Parkin and N. Devlin (2015), "Inter-provider Comparison of Patient Reported Outcomes: Developing and Adjustment to Account for Differences in Patient Case Mix", *Health Economics*, Vol. 24(1):41-54, http://dx.doi.org/10.1002/hec.2999.
[REF-62]	Calvert, M., D. O'Connor and E. Basch (2019), "Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes", *Nature Reviews Drug Discovery*, http://dx.doi.org/10.1038/d41573-019-00088-7.
[REF-63]	McLeod, S. A. (2019). Likert Scale Definition, Examples and Analysis. Simply Psychology. https://www.simplypsychology.org/likert-scale.html.
[REF-64]	Ayearst, L. E., Flett, G. L., & Hewitt, P. L. (2012). Where is multidimensional perfectionism in DSM-5? A question posed to the DSM-5 personality and personality disorders work group. *Personality Disorders: Theory, Research, and Treatment,* 3(4), 458–469. https://doi.org/10.1037/a0026354.
[REF-65]	Thayaparan, A. J., & Mahdi, E. (2013). The Patient Satisfaction Questionnaire Short Form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. *Med educ online,* 18, 21747. https://doi.org/10.3402/meo.v18i0.21747.
[REF-66]	Neilson LJ, Patterson J, von Wagner C, et al. (2020) *Frontline Gastroenterology* 11:209–217.doi:10.1136/flgastro-2019-101321
[REF-67]	Laschinger, H.S, McGillis L, Pedersen, C; Almost J. (2005) A Psychometric Analysis of the Patient Satisfaction With Nursing Care Quality Questionnaire, *J Nurs Care Quality* 20(3):220-230
[REF-68]	Torbjørnsen, A., Småstuen, M. C., Jenum, A. K., Årsand, E., & Ribu, L. (2018). The Service User Technology Acceptability Questionnaire: Psychometric Evaluation of the Norwegian Version. *JMIR human factors*, 5(4), e10255. https://doi.org/10.2196/10255.
[REF-69]	Brook J. (2013) SUS: A Retrospective. *J Usability Studies* 8(2):29-40 JUS Journal Template (uxpajournal.org)
[REF-70]	Hinderks A., Schrepp M, Domínguez Mayo MJ, Escalona MJ, Thomaschewski J (2019), Developing a UX KPI based on the user experience questionnaire, *Comp Stand Interfaces*, 65:38-44,

	https://doi.org/10.1016/j.csi.2019.01.007.
[REF-71]	(OECD 2019, Measuring What Matters: The Patient-Reported Indicator Surveys, Measuring-what-matters-the-Patient-Reported-Indicator-Surveys.pdf (oecd.org).
[REF-72]	Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. (2009) Measuring the quality of hospital care: an inventory of indicators. *Intern Med J.* 39(6):352-60. doi: 10.1111/j.1445-5994.2009.01961.x. PMID: 19323697
[REF-73]	Hafez, R., ed. (2020) Measuring Health System Efficiency in Low- and Middle-Income Countries: A Resource Guide. Joint Learning Network for Universal Health Coverage. Available at: Measuring Health System Efficiency in Low- and Middle-Income Countries: A Resource Guide (jointlearningnetwork.org).
[REF-74]	Häkkinen U, Iversen T, Peltola M, Seppälä TT. Malmivaara A, Belicza É, Heijink R, Fattore G, Numerato D, Medin E, Rehnberg C. 2013. Health care performance comparison using a disease-based approach: the EuroHOPE project. *Health Policy* 112 (1-2):100-9.
[REF-75]	Moger TA, Peltola M. (2014) Risk adjustment of health-care performance measures in a multinational register-based study: A pragmatic approach to a complicated topic. *SAGE Open Med*. 8;2:2050312114526589. doi: 10.1177/2050312114526589. PMID: 26770715; PMCID: PMC4607195.
[REF-76]	Häkkinen, U., Iversen, T., Peltola, M., Rehnberg, C., Seppälä, T. T., and On Behalf of the EuroHOPE Study Group (2015) Towards Explaining International Differences in Health Care Performance: Results of the EuroHOPE Project, *Health Econ*., doi: 10.1002/hec.3282.
[REF-77]	Aluko P, Graybill E, Craig D, Henderson C, Drummond M, Wilson ECF, Robalino S, Vale L; on behalf of the Campbell and Cochrane Economics Methods Group. (2021) Chapter 20: Economic evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021).* Cochrane, 2021. Available from www.training.cochrane.org/handbook
[REF-78]	Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programme. Third edition*. Oxford: Oxford University Press.
[REF-79]	Hunter, R. & Shearer, J. (2013) Cost-consequence analysis – an underused method of economic evaluation. Research Design Service. Accessed at http://www.rds-london.nihr.ac.uk/Howto-design-a-study-find-funding/Health-economics/Cost-consequences-analysis.aspx
[REF-80]	Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2016-10). Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, UK: *Oxford University Press.* Retrieved 15 Dec. 2021, from https://oxfordmedicine.com/view/10.1093/med/9780198725923.001.0001/med-9780198725923
[REF-81]	Mauskopf JA, Paul JE, Grant DM, Stergachis A. (1998) The role of cost-consequence analysis in healthcare decision-making. *Pharmacoeconomics* 13(3):277-88. doi: 10.2165/00019053-199813030-00002. PMID: 10178653.
[REF-82]	Javier Sedano Franco, Miguel Portal García, Alejandro Hernandez Arauzo, José Ramón Villar Flecha, Jorge Puente Peinador, and Ramiro Varela Arias. (2013) Sistema inteligente de recarga de vehículos eléctricos: diseño y operación. DYNA

	INGENIERIA E INDUSTRIA, 88:3, 640–647. doi: 10.6036/5788. URL https://doi.org/10.6036/5788.
[REF-83]	Alejandro Hernández-Arauzo, Jorge Puente, Ramiro Varela, and Javier Sedano. (2015) Electric vehicle charging under power and balance constraints as dynamic scheduling. *Comp Industr Engin*, 85: 306–315. doi: 10.1016/j.cie.2015.04.002. URL https://doi.org/10.1016/j.cie.2015.04.002.
[REF-84]	Rajan E, Martinez M, Gorospe E, Al Bawardy B, Dobashi A, Mara KC, Hansel SL, Bruining DH, Murray JA, Leggett CL, Nehra V, Iyer PG, Pasha SF, Leighton JA, Shiff AD, Gurudu SR, Raffals LE, Lavey C, Katzka DA, Chen CH. (2020) Prospective multicenter study to evaluate capsule endoscopy competency using a validated assessment tool. *Gastrointest Endosc.* 91(5):1140-1145. doi: 10.1016/j.gie.2019.12.024. Epub 2019 Dec 26. PMID: 31883863.
[REF-85]	De Geer L, Oscarsson A, Engvall J. (2015) Variability in echocardiographic measurements of left ventricular function in septic shock patients. *Cardiovasc Ultrasound*. 15, 13-19. doi: 10.1186/s12947-015-0015-6. PMID: 25880324; PMCID: PMC4399417.
[REF-86]	Munavalli J.R., Boersma H.J., Rao S.V., van Merode G.G. (2021) Real-Time Capacity Management and Patient Flow Optimization in Hospitals Using AI Methods. In: Masmoudi M., Jarboui B., Siarry P. (eds) *Artif Intel Data Mining in Healthcare*. Springer, Cham. https://doi.org/10.1007/978-3-030-45240-7_3.
[REF-87]	Brouwer CL, Dinkla AM, Vandewinckele L, Crijns W, Claessens M, Verellen D, van Elmpt W. (2020 Nov) Machine learning applications in radiation oncology: Current use and needs to support clinical implementation. Phys Imaging Radiat Oncol. 30:16, 144-148. doi: 10.1016/j.phro.2020.11.002. PMID: 33458358; PMCID: PMC7807598
[REF-88]	Guo C, Huang P, Li Y, Dai J. (2020) Accurate method for evaluating the duration of the entire radiotherapy process. *J Appl Clin Med Phys*. 21:9, 252-258. doi: 10.1002/acm2.12959. Epub 2020 Jul 25. PMID: 32710490; PMCID: PMC7497908
[REF-89]	Cave DR, Fleischer DE, Leighton JA, Faigel DO, Heigh RI, Sharma VK, Gostout CJ, Rajan E, Mergener K, Foley A, Lee M, Bhattacharya K. (2008) A multicenter randomized comparison of the Endocapsule and the Pillcam SB. *Gastrointest Endosc.* 68(3):487-94. doi: 10.1016/j.gie.2007.12.037.
[REF-90]	The Eurostat regional yearbook 2021 Edition is an online *Eurostat publication,* also downloadable in PDF format (ISBN: 978-92-76-37762-7, ISSN: 2363-1716, doi: 10.2785/894358, cat. number: KS-HA-21-001-EN-N)