HealthPros

International Training Network for
Healthcare Performance Intelligence Professionals

POLICY GUIDANCE ON ADVANCING THE PERFORMANCE ASSESSMENT OF INTEGRATED HEALTHCARE SYSTEMS

Healthcare Performance Intelligence Series No. 1.4 2022

HealthPros Fellow
Nicolas Larrain

Principal Investigators
Oliver Groene

EU H2020-
Agreement
No. 765141

HealthPros

Series editors

HealthPros Fellows

Erica Barbazza, Amsterdam UMC, University of Amsterdam
Pinar Kara, Aalborg University Hospital
Mekha Mathew, University of Surrey, University of Oxford
Nicolas Larrain, Optimedis AG
Kendall Gilmore, Sant'Anna School of Advanced Studies
Damir Ivanković, Amsterdam UMC, University of Amsterdam
Óscar Brito Fernandes, Corvinus University of Budapest
Sophie Wang, Optimedis AG
Bernardo Meza, University of Surrey, University of Oxford
Claire Willmington, Sant'Anna School of Advanced Studies
Mircha Poldrugovac, Amsterdam UMC, University of Amsterdam
Véronique Bos, Amsterdam UMC, University of Amsterdam

Principal Investigators

Niek Klazinga, Amsterdam UMC, University of Amsterdam
Dionne Kringos, Amsterdam UMC, University of Amsterdam
Jan Mainz, Aalborg University Hospital
Søren Paaske Johnsen, Aalborg University Hospital
Simon de Lusignan, University of Oxford
Fabrizio Carinci, University of Bologna
Oliver Gröne, Optimedis AG
Sabina Nuti, Sant'Anna School of Advanced Studies
Anna Maria Murante, Sant'Anna School of Advanced Studies
László Gulácsi, Corvinus University of Budapest
Petra Baji, Corvinus University of Budapest

Support staff

Frank Groen, Amsterdam UMC, University of Amsterdam
Laurian Jongejan, Amsterdam UMC, University of Amsterdam

EU H2020-Agreement No. 765141

PREFACE

HealthPros is a H2020 Marie Sklodowska-Curie Innovative Training Network for Healthcare Performance Intelligence Professionals under grant agreement No 765141, running from January 2018–April 2022. Healthcare performance intelligence can be defined as a structured approach to acting on health policies, using knowledge and information generated through scientific methods and health data to systematically measure indicators of health system performance. The network set out with the aim to train a first generation of Healthcare Performance Intelligence Professionals (HealthPros Fellows) that can make effective use of available healthcare performance data in countries to improve integrated services delivery, patient engagement, equality in access to healthcare, health outcomes and reduce waste in healthcare.

Since 2018, HealthPros Fellows have completed innovative research and multidisciplinary training in Canada, Denmark, Germany, Hungary, Italy, the Netherlands and the United Kingdom. As part of their training, Fellows also completed secondments at partner organizations as an opportunity to obtain local guidance and conduct applied research.

Throughout the programme, HealthPros Fellows have worked to develop tools and implement methods to streamline healthcare performance measurement, develop and apply performance-based governance mechanisms and optimize the use of healthcare performance intelligence by different end-users. Topics explored through a healthcare performance intelligence lens in their work include: actionability of performance indicators; composite measures; integrated care; corporate governance tools; patient and citizen engagement; nudging; use of routine databases for performance improvement; and, long-term care. As the COVID-19 pandemic paralleled the HealthPros programme, many Fellows and the network at-large, sought opportunities to conduct a number of COVID-19-related studies at pace with the pandemic's changing context.

Outputs of the HealthPros programme have continuously been published as open access studies in international, peer-reviewed journals. Additionally, Fellows have actively contributed to webinars, conferences, the delivery of courses, policy dialogues, direct country support, and media engagements, among other types of dissemination to continuously share new findings throughout the programme.

This **Healthcare Performance Intelligence Series** represents the culmination of key research findings by the network into a collection of reports providing methodological, practical, and policy guidance. Reports in the series are tailored to different audiences, ranging from policy-makers, hospital

managers, clinicians, and the general public. The development of each report in the series has relied on close collaboration across the HealthPros network. The range of topics and resources making up this series includes the following:

- Practical experience with implementing disparity and composite measures in large-scale routine quality improvement work to support transferability to other HC systems (No. 1.2 2022)
- A practical guide towards actionable healthcare performance indicators: Selecting healthcare performance indicators that are fit for purpose and use for various stakeholders (current)
- Policy guidance on advancing the performance assessment of integrated healthcare systems (current)
- Policy guidance on the use of PREMs to improve health system performance (No. 2.2 2022)
- Policy summary report on the value of results-based tools in health care management-Lessons learned from COVID-19 dashboards (No. 2.3 2022)
- Business model for effectively involving patients in the financial decision-making of health insurance funds- A guide to health care insurers on fostering the engagement of citizens based on recent experiences in the Netherlands. (No. 2.4 2022)
- Policy summary report on best practices for linking financial incentives to health care performance at individual health care provider, institutional and regional level- A business case for value-based health care systems based on performance intelligence (No. 2.5 2022)
- Policy recommendations on the role of nudging for health care performance assessment agencies (No. 3.2 2022)

The full series of reports can be found online (https://www.healthpros-h2020.eu/). For questions related to the series or HealthPros network please contact Dionne Kringos, PhD (d.s.kringos@amsterdamumc.nl).

Contents

EU H2020-Agreement No. 765141

Executive Summary

What is the problem?

Many countries are introducing health care reforms to encourage the development of integrated health care systems. For reforms that introduce elements of piloting at a sub-national level, there is a keen interest in evaluating the impact of the pilot. For national or long-term reforms, there is a need to define performance indicators. However, a lack of precise tools and guidelines that consider the unique nature of integrated care systems introduces challenges for healthcare managers and policymakers. These, in turn, may lead to a misappreciation of the effects of integrated systems, an unawareness of performance drivers, and, consequently, of insight to guide health system improvement.

How is it addressed?

We reviewed literature specific to performance assessment of integrated healthcare systems and complemented it with the research findings by the HealthPros network. The objective is to compile key levers and elements to assess integrated care system performance and thus, contribute to the scale-up of integrated healthcare initiatives.

How to measure the performance of integrated healthcare systems?

- Performance assessment designs should incorporate four main elements: defining the system's goal, identifying the systems' value-creating mechanisms, selecting indicators, and transforming the assessment into an improvement plan.
- There is consensus in applying a Donabedian approach for performance assessment (Structure – Process – Outcomes). The outcome level assessment will be related to evaluate performance over the system's goals, while the structure and process levels are related to evaluate the performance of the integrated care value-creating mechanisms.
- The 'triple aim' (population health, patient/carer experience, and healthcare costs) is the primary goal of integrated healthcare systems.
- Integrated healthcare systems' value-creating mechanisms are determined by the actions taken in pursuit of the triple aim.
    - At the structure level, political support, governance systems, stakeholder engagement, information integration, IT infrastructure & capacity, monitoring systems, health system workforce and alignment with national health system objectives are essential value-creating mechanisms. Structural aspects of integrated systems are best coordinated by a single entity, known as a 'regional integrator'.

EU H2020-Agreement No. 765141

- At the process level, value-creating mechanisms can be classified in 6 areas: Health intelligence; Data strategy; Disease Management; Health Promotion; System coordination; and Communication & Dissemination.

- Transforming performance assessment into an improvement plan will be determined by contextual factors and the understanding of the interdependencies of stakeholders with system performance. To comprehend contextual factors, the following elements should be assessed: Population size and composition, the scale of the initiative and its maturity stage, the breath of integration (vertical or horizontal), the level at which the system in question operates (micro, meso or macro), the measurement perspective, the timing of the assessment and comparison standards.

Selecting indicators

- Indicators need to be valid and reliable and assessed according to criteria such as importance, scientific acceptability, usability, and feasibility.
- There is substantial literature gathering and proposing indicators to measure the performance of integrated systems. Based on our review, we compiled a list of indicators, with indicators being classified by source, construct, burden in data collection, time frame to assess impact, ability to be influenced by integrated care, and the Donabedian level of performance assessment. The main constructs in the list are:
  - Outcome level: health outcomes, disease burden (including composite measure such as disability-adjusted life years/years lived with disability or age at onset of inpatient nursing need) and risk factors. Also, patient/carer health related quality of life (PROMs) or experience (PREMs), plus sub-constructs related to a safe, effective, timely, patient-centred, equitable, and efficient care experience. Health costs are measured as costs or healthcare utilization.
  - Process level constructs relate to value-creating mechanisms, such as access to care. care coordination, patient safety, improvement in medication, among many others.
  - Indicators at the structure level include patient demographics and provider characteristics.

- We provide our selection of key indicators considering Donabedian levels, triple aim objectives, value-creating mechanisms, levels of care, time frame to reflect impact, types of data and actionability, importance, scientific acceptability, and feasibility.

Data Strategy

- Performance assessment of integrated health care systems requires several types of data and data sources. Data sources are commonly: Insurance claims, care provider administrative information, (electronic) health records, ad-hoc surveys, patient reported outcomes, patient reported experiences and non-health care related data, such as official statistics on population

demographics, economic and deprivation, and environmental data. The integration across data sources is one of the great challenges of integrated systems. Unique patient identifiers facilitate data integration across multiple sources.

- Integrated healthcare systems strive towards establishing a primary data unit composed of a patient level characterization of medical records, healthcare utilization; healthcare cost, behavioural factors, physiological information, and socio-demo-economic information. Likewise, efforts are made to have periodical data to construct time series and panels.

Challenges

- Due to the complexity of integrated care systems and the challenge to establish clear causal patterns for causes and effects, ambiguity remains on the translation of performance measures into actions for improvement.
- Integrated care systems can rarely be assessed by typically controlled trial designs (RCTs), but rather require synthetic control groups. The science of the construction of such control groups is still developing and in terms of internal validity not yet on par with fully controlled designs.
- Core values of integrated care systems, like patient-centredness or a population approach, may conflict with common evaluation practices (that require the exclusion of some patient groups or Winsorization based on cost thresholds). Further, it is difficult to find suitable performance measures for some value-creating mechanisms of integrated systems.
- Other challenges relate to the novelty of integrated healthcare systems: lack of political coordination and support, financial barriers, and lack of stakeholders' engagement, among others.

1. Introduction

The evolution of integrated healthcare systems is a response to a major health policy challenge: An increasing demand for healthcare, driven by newly available treatments, demographic changes and increasing disease prevalence, together with increasingly pressing budgetary restrictions. The burden of chronic disease in Europe is rapidly increasing whilst populations are ageing [1]. In 2017, an estimated 50 million people in the European Union lived with multiple chronic diseases [2]; this number continues to increase at a very significant upward trend [3]. Combined with an increase in the number and cost of new and more effective treatments, budgetary constraints are becoming more pressing. Recently, countries have made great efforts to address the health and economic crisis caused by the Covid-19 pandemic, which adds even more pressure to health systems budget constraints. Furthermore, health systems have to address the backlog of postponed services and the resulting and inevitable detrimental health consequences caused by the pandemic [4].

Supply-led health systems counteract integration and indeed reinforce care fragmentation. Supply-led health systems are designed such that reimbursement is tied to healthcare services provided or expected, instead of focusing on health outcomes and responsiveness to patients´ needs. Consequently, healthcare is structured around services delivery and hence fragmented between providers [5]. Many health systems are still focused on the artificial separation of different levels of care and on ´reactive´ health care delivery. In reactive health systems, services such as prevention, patient-centred care, disease management, care coordination and cross-provider care organization are not systematically incentivized [5,6] and are often implemented on top of the supply-led service delivery system. The major policy challenges outlined above, however, requires a more ´pro-active´ approach, a service delivery system that is person-centred, promotes health, encourages healthy behaviour and anticipates health needs before they lead to diagnoses, disease progression, hospital admission, surgery or rehabilitation [7]. Further, they are part of the roadmap to achieve high quality and high-value care[8]. Health system fragmentation also contributes to increasing healthcare costs. Hospital, primary, and social care spending are siloed, leaving no incentive to optimize outcomes across care pathways at the patient level and enabling healthcare waste [9].

Integrated healthcare systems are defined by value-based principles and seek to shift the focus of healthcare from a supply-led to an outcomes-led approach [5]. The outcomes-led approach is focused on the pursuit at the population level of the 'Triple aim'[10]. In practice, Integrated care is the approach to healthcare that seeks to overcome care fragmentation by linkage and coordination of providers' services along the continuum of care [11]. Integrated care initiatives create, support, and manage integrated healthcare systems. These systems are characterized by their focus on investment

in prevention and health intelligence, improving system efficiency, quality of life, and avoiding unnecessary costs [11,12]. Integrated healthcare systems can respond to the comprehensive needs of people with chronic illness, patients' personalized care needs and deal with budget constraints the best way imaginable: Keeping the overall population as healthy as possible.

Assessing system performance is essential to guide system improvement [13,14]. However, the availability of specific tools, frameworks and guidance on the performance assessment of integrated healthcare systems is limited [11,15] [16,17]. Previous policy briefs addressed the need and challenge of integrated healthcare systems [2] and there is a need to advance the discussion on the performance assessment of such systems [18]. This report presents the findings of state-of-the-art methods in literature and the HealthPros network [19] research in the topic of performance assessment, specific to integrated healthcare systems.

2. Methods

The objective of this report is to compile learnings for policy recommendations in integrated healthcare systems performance assessment. For this purpose, we systematically reviewed published literature examining the performance of integrated healthcare systems. First, we searched PubMed (Boolean terms in appendix) from 2015 onwards (5 articles). Then, we used a snowballing approach starting from two critical reports on the topic: The 'Tools and Methodologies to assess integrated care in Europe' report by the Expert Group on Health Systems Performance Assessment [11] and 'A Guide to Measuring the Triple Aim: Population Health, Experience of Care, and Per Capita Cost' by Matthew Stiefel[15] (7 articles). Finally, we added the learnings suitable for policy recommendations from 7 more articles related to the HealthPros network [19]. Appendix 1 summarizes the reviewed literature and classifies it into two types. The first corresponds to empirical research based on integrated systems performance assessment cases and theory constructions based on expertise and literature revision. The second type are literature reviews of several integrated care cases.

3. Findings

This section outlines the key elements of performance assessment frameworks found in literature, complemented with the findings of the HealthPros network.

3.1 Conceptual models for performance assessment

While not all frameworks for performance assessment are conceptually based on theory, Levesque and Sunderland [20] highlight the advantages of doing so. Using conceptual models to understand the

theory of change will increase the assessment's possibility of triggering the right actions for system improvement. Having a solid theoretical base for performance assessment can, for example, be summarized using a simplified logic model (Box 1) which facilitates:

- Understanding if the current inputs make the assessment feasible.
- Creating measures and measurement techniques to signal the underlying constructs that are crucial for the system.
- And finally, appraising if the expected outputs will produce assessments in accurate context and able to guide the expected system improvement.

The Donabedian approach to performance assessment is established across literature [11,15,21] and is defined by measuring performance at three levels: Structure, Process, and Outcomes. The premise is that 'good structure increases the likelihood of good process, and good process increases the likelihood of good outcome'[22].

It is essential to distinguish between assessing the performance of an integrated system and assessing the level of integration of a health system [11]. While this report focuses on the former, the level of integration of an integrated healthcare system can be seen as a structural and process level element in the performance assessment design. Some authors highlight studying the level of integration as an essential element in performance assessment [11,21]. We recommend reviewing other European projects such as the SCIROCCO [23] and SELFIE [24] programs as approaches of health system integration assessments.

Box 1. Logic Models

A logic model is an overview of a program's resources, planned or implemented activities that maps the causal pathways to achieving the intended results. Most commonly, it considers resources, activities, outputs, outcomes and short to long term goals. By following a logical sequence, one can better understand how the individual components and activities may lead to the desired program goals. Benefits of logic models can be derived both from the development process and the end product [25]. An example of a logic model is given in Table 1.

TABLE 1. EXAMPLE OF A HIGH-LEVEL LOGIC MODEL TO GUIDE THE IMPLEMENTATION OF AN INTEGRATED CARE INITIATIVE

Inputs	Activities		Results		
	Activities	Indicators	Short-term (1 – 3 years)	Medium term (3 – 5 years)	Long-term (> 10 years)
- Support from regional healthcare network, national or regional policies, and local administration. - Establishment of the regional integrator organizations to define governance and ac-countability responsibilities. - Stakeholder engagement: Integrated network of healthcare providers, health insurances, patients' association, and other health related organizations (Schools, pharmacies, etc.) - Capacities, characteristics, and goals of a clearly defined population for which the integrated system is accountable. - Incentives scheme contract in support of	Health intelligence: - Risk stratification, Needs assessment, Impact evaluation, Process evaluation, Feedback & auditing, Decision support, Health intervention evaluation & recommendations, Cost assessment, Utilization of care assessment. Data Strategy: - Data infrastructure & IT, Data governance, Data collection Data access & distribution, Data regulation. Disease Management: - Special care programs: case management, self-management interventions - Shared decision-making, Patient focus/ empowerment/ centeredness Health Promotion: - Public health campaigns, Health	- Providers reporting useful feedback & decision support - Information integration as evaluated by providers - PREMS PROMs - Medication adherence after heart failure - Patient activation(survey)	Successful implementation the regional integrator Enlightened and activated patients Reaching vulnerable target groups Provider compliance with health promotion/ integrated care standards Adequate utilization and transfer behaviour. (Rural/disadvantaged) regions attracting health care providers (initially: marketing advantage) Health as a core element of community identity Attractive working conditions and collaboration opportunities for service providers	Achieving the quadruple Aims: • Reduction of per capita health expenditure • Improving patient satisfaction/experience • Improving population health • Improving the working conditions of service providers Academic/business recognition reflected on publications, certifications.	Reduced disease burden: incidence and/or prevalence of major chronic conditions or delayed onset of disease Delayed nursing need amongst the elderly (compression of morbidity) Strengthening the sustainability of a healthcare system geared to patient needs and expectations

integration of care (e.g., Shared savings, value-based payment. - Financing: Investment from public or private grants or loans. - IT infrastructure to support an integrated data strategy. - Healthcare workforce.	promotion interventions (e.g., health promotion discounts). System coordination, collaboration, and cooperation: - Improve stakeholders' involvement, Care pathway design, Alignment of incentives. Alignment of protocols and guidelines, Promotion of provider, network teamwork, Constant revision & promotion of best practices, Network and population approach to patient care, Innovative supply concepts Communication, dissemination, and exploitation: - Benchmarking, Market access, Knowledge transfer and exchange	- % Provider compliance with health promotion - Vaccination coverage - Healthy lifestyles (smoking%, alcohol dependency%) - Mean age at the time of death; years of potential life lost or gained; premature death; mortality and survival time. - Mean age at the onset of long-term nursing need - Quality of multi-professional teams (survey). - Reducing ambulatory care sensitive hospital admissions - Waiting times for urgent treatment - Patient-level cost savings	Reduction in adverse events (Chronic conditions)		
Assumptions: • Successful fundraising (start-up invest for the regions) • Broad support from the relevant stakeholders • Patients enrolled in the integrated initiative at a high rate, reaching 25% of the population in year 5.			**External factors affecting implementation:** • Stable general health policy conditions • Acceptance of regional integrator function • Development of an improved risk structure equalization between the system payors that positively includes prevention efforts.		

3.2. Defining and contextualizing system goals

The starting point for a performance assessment of integrated health care systems is to define the main goal or goals the system seeks to achieve. At the high level, integrated healthcare systems strive towards three main goals (the 'Triple aim') [11,15]: Improving population health, improving the patient experience of care, and reducing health cost growth. Later, 'improving carer experience of care' has been added, creating the 'Quadruple aim'. In line with integrated care value-based approach, focusing on measuring the performance these multiple aims simultaneously can be understood as a measurement of the system's value[15]: focusing on healthcare costs will provide an assessment from a budget perspective, carer and patient experience will aid assessing the relationship between system costs and operations (efficiency) while focusing on population health helps gauging the relationship between system costs, processes, and its main result, or in other words, a measure of effectiveness.

At the top of the Donabedian approach, outcome measurements focus on the gains of the integrated healthcare system towards the triple aim [15]. Consequently, the threefold objective offers the high-level underlying constructs for performance measurement at this level. Some authors highlight that the triple aim can provide constructs at both the outcome and process levels of performance assessment. An example is given by Stiefel & Nolan [15] when considering smoking status (indicator for the construct of healthy lifestyles) as being directly associated with the triple aim (population health) at the process level of assessment. However, other authors consider that constructs such as healthy lifestyles are at the outcome level of assessment [26,27], making the appropriate classification determined by the system's context. Literature coincides in using the triple aim to guide high-level constructs for performance assessment.

Transforming the high-level system goals into specific objectives will depend on the contextualization of the integrated system in question [15]. Therefore, having a well-defined population and the initiative's scale (e.g., local, regional or national) will be essential for setting specific objectives and contextualizing system performance overall [8,15,28]. Furthermore, integrated healthcare systems are evolving and can be assessed at different stages of maturity. Finally, alignment with country or international healthcare goals, values, and needs is crucial. A suitable alignment will facilitate the political support needed to transition into continuing and scaling up integrated healthcare.

3.3. Understanding value creating-mechanisms

Understanding the value-creating mechanisms of integrated health care systems helps to define what to measure when assessing performance at the structure and process levels [20]. Integrated systems' value-creating mechanisms are the interventions, focus areas and activities that systems pursue to work towards the triple aim. Once the value-creating mechanisms of integrated systems are known, their implementation can be assessed (type, number, range, scope) and evaluated to what extent they contribute to achieving the overall aims.

An overview of these value-creating mechanisms can be gained from the operations of leading integrated healthcare systems. The World Innovation Summit for Health (WISH) Accountable Care Forum report [5,29] selected four initiatives that fit this definition as they focus on established integrated healthcare systems accountable to a clearly defined population.

3.3.1 Structure

Value-creating mechanisms at the structure level provide the essential elements for the correct functioning of integrated systems. First, the complex transition to integrated care needs political support, including the willingness to change the organizational and structural face of the healthcare system [11]. Second, well-defined and robust governance mechanisms are essential to determine accountability and facilitate the interconnections between stakeholders [11,30]. Ideally, this value creating-mechanism is reflected in a facilitator for integrated care [8,21], also called a 'regional integrator' [10]. Third, stakeholder engagement is essential for creating an integrated and accountable health system. An accountable system will promote cooperation, coordination, collaboration, and trust [11], all critical values in integrated care. Fourth, as we mentioned before, a defined population, scale, and alignment with national goals. Fifth, system-wide shared goals and incentives schemes, as all the elements of the integrated system must be pushing towards the same objectives [11,21].

Essential structural elements can vary in levels of maturity or implementation, such as the level of information integration (ideally based on Electronic Health Records), including health, healthcare utilization, behavioural, physiological, socio-economic, and costing data. There is also a need for IT infrastructure, capacity, and intelligence [11]. Moreover, a continuous and effective monitoring system that can detect, correct, and improve system processes and performance [11]. Finally, an adequate capable, and continuously learning health system workforce [11,21]. The first five mechanisms can be better assessed with a checklist, while the rest are more suitable for an ad-hoc process evaluation [31].

3.3.2 Process

Value-creating mechanisms at the process level of assessment correspond to the actions of integrated healthcare systems. Two elements will define what underlying constructs better apply to the context. First, the level of integration of the system, which can be either horizontal (at the same level of care) or vertical (across levels of care) [21,32]. Second, the level at which the system in question operates: Micro (patient-level), Meso (organizational context) and/or macro (financing and policy context) [11,32]. These mechanisms can be compiled and classified into six work areas:

1. **Health intelligence:** Including actions towards risk stratification, needs assessment, impact evaluation, process evaluation, feedback & auditing, decision support tools, health applications evaluation and recommendation, per-patient cost, and utilization of care, among others.
2. **Data Strategy:** Considering the following vital elements: Data infrastructure, Data governance, Data collection, Data access & distribution, Data regulation.
3. **Disease Management:** Including interventions such as special care programs – case management, self-management interventions, shared decision-making interventions, patient focus/empowerment/patient centredness.
4. **Health Promotion:** With actions such as public health campaigns and health promotion interventions (health promotion discounts).
5. **System coordination, collaboration, and cooperation:** Taking proactive action to improve stakeholders' involvement, care pathway design, alignment of incentives, alignment of protocols and guidelines, promotion of provider network teamwork, constant revision & promotion of best practices, network approach to patient care.
6. **Communication, dissemination and exploitation:** Emphasizing health and economic results, benchmarking, market access, and knowledge transfer with top-performing health systems.

3.4 Selecting indicators for performance assessment

Indicators should follow quality criteria to ensure that they will generate the expected impact towards system improvement. Many quality criteria have been highlighted for selecting indicators [15,27,33,34]. We have compiled these elements in Appendix 2. Deliverable 1.3 of this series, "A practical guide towards actionable healthcare performance indicators." expands on the criteria indicators for performance measurements should be selected. Indicators actionability is highlighted and expanded into two key concepts. Fitness for purpose, understanding the extent to which an

indicator serves an intended decision-making function, that is, a task or specific use and; Fitness for use, referring to the potential for an indicator to get the right information into the right hands at the right time.

Further, we analysed the selection criteria used by the National Quality Forum: importance, scientific acceptability, usability, and feasibility. An important indicator can be influenced through care integration. In other words, particular emphasis should be given to selecting indicators that will significantly vary because of the actions of the integrated initiative. Scientific acceptability is related to the importance of the robustness of the measures [11] and their comparability and reproductivity. Usability relates to prioritizing the indicators' fit for use, or in other words, how useful they are for guiding system improvement and triggering action. Also, it relates to both the importance of indicators to be simple to understand and easy to measure [11]. Finally, feasibility urges us to consider if the selected indicators can be constructed with the current inputs, data and how resource-intensive (including time) the assessment will be [27].

In addition, we can highlight two more elements that we deem essential for selecting performance indicators of integrated care and are not directly included in the compilation of elements in the literature. These are the time frame for signalling an effect and the relation of the indicator to economic outputs (both concepts related to indicator actionability). The actions of integrated systems have different expected timeframes to influence system performance, hence indicators should be selected according to the expected timeframe. Likewise, the actions of integrated systems are proactive, and as such they need financing and up-front investment. Therefore, financial indicators are critical to monitor the pathway to financial impact. There is substantial literature gathering and proposing indicators to measure the performance of integrated systems[11,15]. Optimity Advisors[35] compiled a list of indicators using the resources used in the expert group for health system performance assessment of the EU[11]. We have complemented the list with indicators in the reviewed literature and classified them by source, construct, data source, the burden in data collection, time frame to assess impact, the ability to be influenced by integrated care, and the type by the Donabedian approach to performance assessment (Appendix 3). While most classification categories are self-explanatory, 'Construct' refers to the underlying construct of either the systems goals or value-creating mechanisms of integrated care and indicates the concept the indicator is measuring.

At the outcome level, population health is measured with indicators in the constructs of health outcomes (mortality, health status), disease burden (at-risk population, incidence/prevalence), and behavioural and physiological factors (risky behaviours). Patient/carer experience is measured mainly

directly as patient/carer experience, plus sub-constructs related to safe, effective, timely, patient-centred, equitable, and efficient care experience [14]. Healthcare costs are measured directly as healthcare costs, highlighting the indicator 'per capita cost of care', and healthcare utilization. We highlight the need for having all supply, demand, and intermediary (insurances) perspectives in the cost assessment [15].

At the process level, the constructs relate to the value-creating mechanisms exposed in section 3.3. However, it is important to notice that some value-creating mechanisms are not considered in the compiled list of indicators. In the same line, few indicators refer to the structure level, a plausible finding given the remarks in section 3.3. The few existent constructs at the structure level are related to the demographics of the patient population.

Identifying a set of indicators covering all relevant assessment areas and complying with the quality characteristics of an indicator is a challenging task. The expert group for health system performance assessment of the EU selected a list of 19 indicators with this purpose (they can be found in the compilation of indicators in Appendix 3). These indicators contemplate constructs that implicate a wide range of providers and can be heavily influenced by the system's integration, hence deemed highly important. They are focused on mortality, readmission, and follow-up (medication) after hospital care. Mortality is related to population health, readmission to care efficiency, medication follow-up to prevention, patient-centeredness, and care coordination (secondary-primary). However, it is heavily focused on outcome and process measures and leans towards hospital care. Also, they are all short to mid-range in time frame and do not refer to different data types, though various data sources are considered. These shortcomings demonstrate the difficulty of the task. Taking into consideration these issues we present a comprehensive and manageable list of key indicators that serve as a basic set for assessing performance of an integrated healthcare system (Figure 1).

Figure 1. Key set of performance indicators for evaluating integrated healthcare systems.

System goals, Triple aim	Value-creating mechanisms	Results can be observed Short-term indicators (<2 years)	Medium-term indicators (2-3 years)	Long-term indicators (5+ years)
Population Health				**Mortality outcomes: Mean age at the time of death; years of potential life lost or gained; premature death; mortality and survival time.**
		Increase in patient activation		**Mean age at the onset of long-term nursing need**
	Health Promotion	*% Provider compliance with health promotion/integrated care standards*	**Vaccination coverage** *Changes towards healthy lifestyles*	**Disease Burden: Incidence (yearly rate of onset, avg. age of onset) and/or prevalence of major chronic conditions**
Patient/Carer experience	System coordination	**Reduction in adverse events (Chronic conditions)** *Quality of multi-professional teams as assessed by surveys*	**Reducing ambulatory care sensitive hospital admissions** **Waiting times for urgent treatment (esp. cancer, severe mental health access, elective treatment)**	**Mortality or all-cause readmission within 30 and 365 days after discharge**
	Disease Management	**Medication adherence after heart failure**	*Patient experience of involvement in personalized care plan development*	**Mortality or disease-specific readmission within 30 and 365 days after discharge**
		Percentage of patients that have received help to stop smoking after an AMI		
	Health intelligence		*Providers reporting useful feedback & decision support*	
	Data strategy	*Information integration as evaluated by providers*		
	Communication & Dissemination		*Recognition, acceptability and impact of health communication efforts in the community* *Academic/business recognition reflected on publications, certifications*	
Financial performance			**Patient-level cost savings over the whole continuum of care relative to the costs normally expected**	

Legend

Coding	Donabedian levels	N°
	Structure	2
	Process	10
	Outcome	8

Coding	Type of data	N°
Italic	ad-hoc surveys	7
Bold	Claims/Administrative	13

EU H2020-Agreement No. 765141

3.5 Understanding and contextualizing performance assessment results

Performance assessment does not end when the performance measures are constructed. A plan to contextualize results and trigger the appropriate actions for improving performance is needed. Even though a challenge, laying out the interdependencies of stakeholders and their relation to system performance will guide the transformation of data into action.

Understanding the timing of the assessment and linking it to the stage of the integrated systems is essential to contextualize results. Indicators will vary in the time delay they present for signalling the effect of an intervention, and this should be considered when assessing performance [8,11].

Moreover, current literature makes efforts in creating assessment frameworks with accurate performance indicators for the structural components, value-creating mechanisms, and objectives unique to integrated systems. Nevertheless, they provide limited information on evaluation designs and on how to enhance causal inference [36], contributing to evaluations with severe limitations [28]. Two methods for evaluation are commonly applied. One is understanding performance in time, for which time series data or, panel data is needed [15]. The second is to compare a system's performance to benchmarks [15,37] and/or control groups. Benchmarking can improve performance through reciprocal altruism mechanisms [37]. Control group comparisons on the other hand will guide system improvement by informing about an intervention's impact over system performance.

Typically, experiments, quasi-experiments, and comparisons of standardized assessments are used for this purpose. Standardized assessments are very useful to contextualize results, but they are often conducted as observational studies and can't determine causality over the outcomes with certainty. Further, both from an ethical and a pragmatic perspective, it is virtually impossible to conduct RCTs [38]. Quasi-experimental designs combine rigour and flexibility adequate for evaluating impact of integrated systems, and researchers have used them before for this purpose [28,36]. Table 2 presents a high-level overview of the most common quasi-experimental methods in healthcare impact evaluation. These designs use statistical methods to create a counterfactual that captures what would have been the outcomes if the programme/policy had not been implemented [40]. Depending on the data at hand, the estimation methods can be ex-post single difference or double difference (also known as difference-in-differences or DID). The difference between the approaches is related to the method used for minimizing the counterfactual's risk of bias, or in other words, be as certain as possible that the only difference between groups is the effect of the intervention being evaluated.

TABLE 2. OVERVIEW OF COMMON QUASI-EXPERIMENTAL DESIGNS.

Method	Overview
Regression discontinuity [41]	The design is defined by identifying a mechanism that defines the assignment of the treatment. Once we control by the mechanism (assuming its level is uncontrollable by participants) the assignment of intervention is nearly as good as randomization for the participants around the threshold, as the only difference between them is the intervention. The method requires the intervention to be assigned by a clear qualitative threshold on a (running) variable, such as cut scores for university selection.
Natural Experiments [42]	The method requires a natural treatment assignation exogenous to the variable of interest or its determinants. As such, it can be thought of as a random assignation. A random treatment assignation will secure that the treated and untreated portions of the sample (if big enough) will have, on average, comparable characteristics and, hence, their difference in the variable of interest can be attributed to the treatment.
Instrumental variables [43]/ sample selection (Heckman) models [44]	These methods are regression based and take care of selection bias by introducing a statistical correction to the non-random treatment assignation. The correction in instrumental variables is related to finding a different variable (instrument) that is related to the conflicting variable but not to the outcome of interest. In Heckman models, the probability of receiving treatment is modelled for each observation and included in the estimation as a corrector in the treatment regressor.
Matching [40,45]	Matching methods rely on observed characteristics to find statistical twins(matches) to the treated units among untreated units and, by doing so, constructing a comparable control group.
- **Exact**	Exact matching refers to finding matches to each of the units in the treated sample so that they have the same value in the observed covariates.
- **Propensity Score Matching**	With this method, treated units are not matched on observable characteristics but on the likelihood that the individual will be assigned to treatment (propensity score), in time determined by observable characteristics.
- **Mahalanobis metric matching**	A common alternative to propensity score matching when there are few covariates (<8), the method matches treated units by comparing the Mahalanobis metric. This metric is a multi-dimensional generalization of the idea of measuring how many standard deviations away an observed covariate value is from the mean of the distribution of said covariate.
- **Genetic Matching [46]**	Using an evolutionary algorithm, the method generalizes previously seen matching approaches to find the control group that optimizes the post-matching observed covariate balance.
Entropy Balancing [47]	It is a multivariate reweighting method to produce balanced samples. In a pre-processing procedure, the methodology calculates control unit weights so that the reweighted control group satisfies a pre-set of balance conditions that are imposed on the sample moments of the observed covariate distributions. The optimization problem will search for the set of weights that satisfies the balance constraints and create comparable samples where, on average, the only difference between them is the treatment in question.
Synthetic control groups	Mostly used when there is one or few treated units and many untreated units observed through time. The method consists of creating a control group for the treated unit out of a weighted combination of the untreated units.

EU H2020-Agreement No. 765141

The assessment plan should be contextualized to local/regional factors to ensure that the main objectives of the assessment are met. Further. efforts should be made to understand how performance measures will be translated into action for improvement. The selection of indicators based on the criterion ´actionability´ is not sufficient, rather the roadmap to trigger actions should be laid out in the assessment plan. For example, Bos et al. [30] encounters limited actionability of performance intelligence in a Dutch integrated healthcare system due to a misalignment between care providers and their respective catchment areas. Considering which stakeholders are interacting with the underlying constructs, and from which perspectives, will help to understand the roadmap to transform the assessment into action [8]. Further, there is a need for empirically tested theoretical models, and more work is needed to clarify relationships between underlying concepts and how stakeholders and outcomes are related to them [48].

3.6 Building on an integrated data strategy

Data is necessary to construct the measures that will assess system performance. At the same time, it is essential to measure the performance of the system regarding its data strategy. Data strategy is the structured set of choices an organization takes regarding how data is identified; stored; provided; processed, and governed [49]. Integrated systems seek to construct a primary data unit containing a patient-level characterization of clinical records, healthcare utilization, healthcare cost, behavioural factors, physiological information, and socio-demo-economic information [50]. Likewise, periodic data to construct time series and panels are a core elements of successful data strategies [15].

Given the wide range of goals and value-creating mechanisms in integrated systems, there is a need for several types of data. Health service administrative data, clinical data, population health data (including health status and behavioural and physiological factors), experience data, socio-demo-economic data, and claims or costing data are identified in the reviewed literature.

Health services administrative data, derived from claims data, is mainly used to construct information about healthcare utilization, but can be used to construct various relevant outcome measures, too. Clinical data comes from patient records, where Electronic Health Records provide an immense advantage for a successful data strategy. Population health data is constructed from different sources. Health status data comes from patient records and patient recorded outcomes (PROMs). Behavioural factors data will typically come from specific surveys or questionnaires. Patient recorded experiences (PREMs) are the primary source for experience data. Public records and specific questionnaires provide socio-demo-economic data. Finally, claims or costing data comes from the

health system payors (national health funds, health insurance companies) and care provider administration. Claims data can be comprehensive and can even partially substitute clinical and administrative data to construct patient records including health status and healthcare utilization information. Typically, this substitution is closely related to morbidity and healthcare use information but lacks behavioural and physiological information.

The integration across data sources is one of the significant challenges of integrated systems. Unique patient identifiers facilitate said integration. Ultimately, success in data integration will be determined by stakeholders' engagement, alignment of political and business perspectives, and a high level of system integration overall.

3.7 Addressing complexity through methodological advances

The unique nature of integrated healthcare systems presents challenges for performance assessment. While enhancing the interconnections of the system's stakeholders at various levels, integrated care increases the complexity of an already complex system [11,51]. Under these circumstances, understanding the causal effects between the system actions and the observed results becomes more challenging, and stakeholders' accountability for results (good or bad) might get lost in the system interdependency [52,53]. The solution is to complement the monitoring of performance indicators with models capable of integrating complexity when assessing performance [42]. Simulation models are an excellent example of such complements [12].

Quasi-experiments, the evaluation design better equipped to evaluate integrated healthcare systems, have the challenge of identifying an appropriate counterfactual. Separating the evaluation according to the types of effect and sub-sampling population that fits partitioned evaluations can facilitate this task and generate more accurate assessments [54]. However, common methods for the creation of control groups using statistical analysis (matching methods) have a shortcoming that is particularly detrimental to integrated healthcare systems. Cases of high leverage or low prevalence (particularly in small to medium-size systems) will typically be excluded from the analysis. This practice is called 'Winsorization', and it is harmless when the effect of an intervention is constant for the affected population. However, the population approach of integrated systems and their accountability for the full spectrum of care, pushes the system to have differentiated interventions for all its patients. In this context we recommend to utilize methods that don't exclude participants, such as the novel Entropy Balancing method [39,47].

Innovative assessment methods can be used to assess integrated care value creating mechanisms that are difficult to measure. IT-based measures, EHR-based measures, and Social Network Analysis are examples of these methods [8]. Cooperation between care providers, for example, is one core concept that is overlooked in the existing frameworks of integrated care performance assessment and that has been successfully assessed by researchers of the HealthPros network with a methodology based on Social Network Analysis [55].

Data intelligence constitutes an essential part of the integrated care approach. To address the uncertainty about how data should be used in integrated healthcare systems and contribute in determining the core elements of a successful data strategy, researches at the HealthPros Network reviewed the operations of 4 leading cases of integrated care in regards to their data strategy[50]. The main findings recommend having a clearly defined and explicit data strategy, comprehending several data types and sources, appropriate infrastructure to provide the right data to the different users so that they can transform data into actions that actively generate value, a culture of constant innovation, and unambiguous rules and responsibilities for all the data strategy stakeholders.

Integrated healthcare novelty and relatively new and growing appreciation creates challenges that are important to attend. Political support and avoiding contradicting policies will be essential for the initiative's functioning. Appropriate financing will secure the necessary participation of both users and care providers. Finally, a well-resourced system in terms of inputs, services and human capacity, together with engaged stakeholders will be crucial for the success of integrated systems [21].

3.8 Summary of recommendations

1. Design a performance assessment plan that is conceptually based on theory. Logic models are a good tool for this purpose. The plan should describe and define the causal link between inputs, activities, outputs and short-, mid- and long-term objectives.
2. Contextualizing the integrated system will be essential for an accurate assessment. Crucial topics to pay attention comprehend a well-defined population and scale, the stage of development of the system, the alignment with country and international goals, the level of integration and understanding at what level of the health system integrated care is occurring.
3. The assessment plan should assess performance at structure, process, and outcome levels.
4. Define the triple aim as integrated system's objective. As such, the triple aim provides the primary constructs that need to be assessed at the outcome level.

5. Understand the system's value creating mechanisms. They will provide the constructs to measure performance at the process and structure levels.
6. Select indicators that are important, scientifically acceptable, usable, and feasible. Further, we recommend paying special attention to indicator's actionability, time frame to signal effects and relation to economic outputs. We provide a chart with our selection of key indicators for performance assessment of integrated healthcare systems (Figure 1).
7. Benchmarking, evaluating impact and understanding performance evolution are important to contextualize performance assessment results and guide system improvement.
8. Complete the performance assessment plan with a roadmap that guides system improvement. Understanding the interdependencies and responsibilities of system's stakeholders will help in designing an improvement plan.
9. Define your data strategy. This should include: The data that you need, the sources that will provide it, how will data be stored and processed, what analysis will you execute and, the rules, owners, and responsibilities for the data users. The gold standard is a primary data unit considering a patient-level characterization of clinical records, healthcare utilization, healthcare cost, behavioural factors, physiological information, and socio-demo-economic information. Further, having periodical data and a plan to transform data into action is crucial for system improvement.
10. Researchers should consider the unique nature of integrated healthcare when evaluating these systems' performance. Emphasis should be given to the system complexity when understanding causal paths of system interventions. Likewise, the effects of perspectives like the population health approach over methodologies for constructing control groups should be considered when evaluating impact. Finally, innovative IT based methods can be used to create indicators for value-creating mechanisms that are hard to measure.

EU H2020-Agreement No. 765141

4. References

1. Organisation for Economic Co-operation and Development, Europeiska unionen. Health at a glance: Europe 2020 : state of health in the EU cycle. 2020.

2. Rijken M, Struckmann V, van der Heide I, Hujala A, Barbabella F, van Ginneken E, et al. Policy brief (European Observatory on Health Systems and Policies. Online). European Observatory on Health Systems and Policies; 2016. (Health systems and policy analysis).

3. Souza DLB, Oliveras-Fabregas A, Minobes-Molina E, de Camargo Cancela M, Galbany-Estragués P, Jerez-Roig J. Trends of multimorbidity in 15 European countries: a population-based study in community-dwelling adults aged 50 and over. BMC Public Health [Internet]. 2021 Dec [cited 2022 Jan 4];21(1):76. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-10084-x

4. WHO RO for E. Spending on health in Europe: entering a new era. Copenhagen: WHO: WHO Regional Office for Europe; 2021. Report No.: ISBN 9789289055079.

5. McClellan M, Kent J, Stephen B, Macdonnell M, Thoumi A, Shuttleworth B, et al. Accountable Care: FOCUSING ACCOUNTABILITY ON THE OUTCOMES: Report of the Accountable Care Working Group 2013 [Internet]. World Innovation Summit for Health (WISH), Qatar Foundation; 2013. Available from: https://www.wish.org.qa/wp-content/uploads/2018/01/27425_WISH_Accountable_care_Report_AW-Web.pdf

6. Grilo AM, dos Santos MC, Gomes AI, Rita JS. Promoting Patient-Centered Care in Chronic Disease. In: Sayligil O, editor. Patient Centered Medicine [Internet]. InTech; 2017 [cited 2022 Jan 4]. Available from: http://www.intechopen.com/books/patient-centered-medicine/promoting-patient-centered-care-in-chronic-disease

7. European Commission. “EU integrated approach to non-communicable diseases” [Internet]. 2020. Available from: https://ec.europa.eu/health/non_communicable_diseases/overview_en.

8. McDonald K, Schultz E, Albin L, Pineda N, Lonhart J, Sundaram V, et al. Care Coordination Measures Atlas [Internet]. Rockville, MD.: Agency for Healthcare Research and Quality; 2014. Available from: https://www.ahrq.gov/ncepcr/care/coordination/atlas.html

9. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System: Estimated Costs and Potential for Savings. JAMA [Internet]. 2019 Oct 15 [cited 2021 Feb 12];322(15):1501. Available from: https://jamanetwork.com/journals/jama/fullarticle/2752664

10. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. Health Affairs [Internet]. 2008 May [cited 2018 Sep 10];27(3):759–69. Available from: http://www.healthaffairs.org/doi/10.1377/hlthaff.27.3.759

11. European Commission. Directorate General for Health and Food Safety. Blocks :tools and methodologies to assess integrated care in Europe : report by the Expert Group on Health Systems Performance Assessment. [Internet]. LU: Publications Office; 2017 [cited 2021 Mar 5]. Available from: https://data.europa.eu/doi/10.2875/017891

12. Larrain N, Groene O. Simulation modeling to assess performance of integrated healthcare systems: Literature review to characterize the field and visual aid to guide model selection. Kuo Y-H, editor. PLoS ONE [Internet]. 2021 Jul 9 [cited 2022 Jan 4];16(7):e0254334. Available from: https://dx.plos.org/10.1371/journal.pone.0254334

13. Smith P, editor. Performance measurement for health system improvement: experiences, challenges and prospects. Cambridge ; New York: Cambridge University Press; 2009. 726 p. (The Cambridge health economics, policy and management series).

14. Papanicolas I, European Observatory on Health Systems and Policies, editors. Health system performance comparison: an agenda for policy, information and research. Maidenhead, Berkshire, England: Open University Press; 2013. 384 p. (European observatory on health systems and policies series).

15. Stiefel M, Nolan K. A Guide to Measuring the Triple Aim: Population Health, Experience of Care, and Per Capita Cost. Cambridge, Massachusetts: Institute for Healthcare Improvement; 2012. (IHI Innovation Series white paper).

16. Dates M, Lennox-Chugani N, Pereira HS, Tedeschi M. Health system performance assessment – Integrated Care Assessment (20157303 HSPA). Brussels: Public Health - European Commission; 2018.

17. Nuño Solinís R, Stein KV. Measuring Integrated Care – The Quest for Disentangling a Gordian Knot. International Journal of Integrated Care [Internet]. 2016 Oct 17 [cited 2021 Mar 5];16(3):18. Available from: http://www.ijic.org/articles/10.5334/ijic.2525/

18. Kaló Z, Zemplényi A, Rutten-van Mölken M, Looman W, Huić M, Tandara Haček R, et al. Development of transferability guidance for integrated care models with special focus on Central and Eastern European countries. Croat Med J. 2020 Jul 5;61(3):252–9.

19. Kringos DS, Groene O, Johnsen SP. Training the First Generation of Health Care Performance Intelligence Professionals in Europe and Canada: Academic Medicine [Internet]. 2019 Jun [cited 2022 Jan 6];94(6):747–8. Available from: http://journals.lww.com/00001888-201906000-00007

20. Levesque J-F, Sutherland K. Combining patient, clinical and system perspectives in assessing performance in healthcare: an integrated measurement framework. BMC Health Serv Res. 2020 Jan 8;20(1):23.

21. Pinter KA, Zhang H, Liu C, Tran B, Chokshi M, Lucerno-Prisno DE, et al. Elements and Performance Indicators of Integrated Healthcare Programmes on Chronic Diseases in Six Countries in the Asia-Pacific Region: A Scoping Review. Int J Integr Care [Internet]. 2021 Feb 8 [cited 2022 Jan 6];21(1):3. Available from: https://www.ijic.org/article/10.5334/ijic.5439/

22. Donabedian A. The quality of care. How can it be assessed? JAMA. 1988 Sep 23;260(12):1743–8.

23. The SCIROCCO project (Scaling Integrated Care in Context) [Internet]. Vol. The project Scirocco is co-funded by the Health Programme of the European Union under Grant Agreement 710033 (CHAFEA). Available from: www.scirocco-project.eu

24. SUSTAINABLE INTEGRATED CARE MODELS FOR MULTI-MORBIDITY DELIVERY, FINANCING AND PERFORMANCE (SELFIE) [Internet]. Vol. An EU-funded Horizon2020 project in PHC-23-2014. Available from: https://www.selfie2020.eu/

25. Porteous NL, Stewart PJ, Sheldrick BJ, Ottawa-Carleton (Ont.), Health Dept. Program evaluation tool kit: a blueprint for public health management. Ottawa: Regional Municipality of Ottawa-Carleton Public Health Depart.; 1997.

26. World Health Organization (WHO). WHO global strategy on people-centred and integrated health services: interim report [Internet]. 2015. Available from: https://apps.who.int/iris/handle/10665/155002

27. Raleigh V, Bardsley M, Wistow G, Wittenberg R, Erens B, Mays N. Integrated care and support Pioneers: indicators for measuring the quality of integrated care. London: Policy Innovation Research Unit, London School of Hygiene & Tropical Medicine; 2014.

28. Comendeiro-Maaløe M, Ridao-López M, Gorgemans S, Bernal-Delgado E. A comparative performance analysis of a renowned public private partnership for health care provision in Spain between 2003 and 2015. Health Policy. 2019 Apr;123(4):412–8.

29. McClellan M, Thoumi A, Udayakumar K, Patel H, Abaou Samra AB. IMPLEMENTING ACCOUNTABLE CARE TO ACHIEVE BETTER HEALTH AT A LOWER COST [Internet]. WISH Accountable Care Forum 2016; 2016. Available from: https://www.wish.org.qa/wp-content/uploads/2018/01/IMPJ4495_WISH_Accountable_Care_Report_WEB-1.pdf

30. Bos VLLC, Klazinga NS, Kringos DS. Improving performance intelligence for governing an integrated health and social care delivery network: a case study on the Amsterdam Noord district. BMC Health Serv Res [Internet]. 2021 Dec [cited 2022 Jan 6];21(1):517. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-021-06558-2

31. Linnan L, Steckler A. Process evaluation for public health interventions and research. 2002;

32. WHO Regional Office for Europe. Integrated care: An Overview [Internet]. World Health Organization; 2016. Available from: https://www.euro.who.int/__data/assets/pdf_file/0005/322475/Integrated-care-models-overview.pdf

33. Pestronk R. Using Metrics to Improve Population Health [Internet]. Prev Chronic Dis; [cited 2022 Mar 1]. (2010;7(4):A70.). Available from: http://www.cdc.gov/pcd/issues/2010/jul/10_0018.htm

34. Bilheimer L. Evaluating metrics to improve population health [Internet]. Prev Chronic Dis; [cited 2022 Mar 1]. Report No.: 2010;7(4):A69. Available from: http://www.cdc.gov/pcd/issues/2010/jul/10_0016.htm.

35. Optimity Advisors. Performance Assessment Framework [Internet]. 2018 [cited 2022 Jan 14]. (EU Third Health Programme [2014-2020]). Available from: https://ec.europa.eu/chafea/health/archives/documents/health/icic-18-santanapereira_en.pdf;

36. Pimperl A, Schulte T, Mühlbacher A, Rosenmöller M, Busse R, Groene O, et al. Evaluating the Impact of an Accountable Care Organization on Population Health: The Quasi-Experimental Design of the German Gesundes Kinzigtal. Population Health Management [Internet]. 2017 Jun [cited 2021 Apr 15];20(3):239–48. Available from: https://www.liebertpub.com/doi/10.1089/pop.2016.0036

37. Bevan G, Evans A, Nuti S. Reputations count: why benchmarking performance is improving health care across the world. HEPL [Internet]. 2019 Apr [cited 2022 Mar 2];14(2):141–61. Available from: https://www.cambridge.org/core/product/identifier/S1744133117000561/type/journal_article

38. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ [Internet]. 1996 May 11 [cited 2022 Jan 27];312(7040):1215–8. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.312.7040.1215

39. Larrain N, Groene O. Improving the evaluation of integrated healthcare systems using entropy balancing. Working Paper. 2022;

40. White H, Sabarwal S, UNICEF. Quasi-Experimental Design and Methods: Methodological Briefs - Impact Evaluation No. 8. (Methodological Briefs no. 8,).

41. Pischke J-S. Regression Discontinuity Design [Internet]. London: London School Of Economics; 2018. Available from: https://econ.lse.ac.uk/staff/spischke/ec533/RD.pdf

42. Petticrew M, Knai C, Thomas J, Rehfuess EA, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision making. BMJ Glob Health [Internet]. 2019 Jan [cited 2021 Mar 4];4(Suppl 1):e000899. Available from: https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2018-000899

43. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidem Drug Safe [Internet]. 2010 Mar 30 [cited 2022 Jan 29];19(6):537–54. Available from: https://onlinelibrary.wiley.com/doi/10.1002/pds.1908

44. Heckman JJ. Sample Selection Bias as a Specification Error. Econometrica [Internet]. 1979 Jan [cited 2022 Jan 29];47(1):153. Available from: https://www.jstor.org/stable/1912352?origin=crossref

45. Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. Statist Sci [Internet]. 2010 Feb 1 [cited 2021 Apr 27];25(1). Available from: https://projecteuclid.org/journals/statistical-science/volume-25/issue-1/Matching-Methods-for-Causal-Inference--A-Review-and-a/10.1214/09-STS313.full

46. Taiwo A. A. Quasi-Experimental Design: Genetic Matching [Internet]. Taiwo A. Ahmed. [cited 2022 Sep 29]. Available from: https://taiwoahmed.com/2020/11/27/genetic-matching-for-causal-inference-in-experimental-and-observational-studies/

47. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Polit anal [Internet]. 2012 [cited 2021 Apr 23];20(1):25–46. Available from: https://www.cambridge.org/core/product/identifier/S1047198700012997/type/journal_article

48. Strandberg-Larsen M, Krasnik A. Measurement of integrated healthcare delivery: a systematic review of methods and future research directions. Int J Integr Care [Internet]. 2009 Feb 4 [cited 2021 Mar 5];9(1). Available from: http://www.ijic.org/article/10.5334/ijic.305/

49. The 5 Essential Components of a Data Strategy. SAS Institute; 2018.

50. Larrain N, Morando V, Groene O. Data strategy for Population Health Management: Review of leading Accountable Care Organizations. Working Paper [Internet]. 2022; Available from: For an unpublished version contact Nicolas Larrain: n.larrain@optimedis.de

51. Edgren L, Barnard K. Complex adaptive systems for management of integrated care. Leadership in Health Services [Internet]. 2012 Jan 27 [cited 2021 Mar 4];25(1):39–51. Available from: https://www.emerald.com/insight/content/doi/10.1108/17511871211198061/full/html

52. Bautista MAC, Nurjono M, Lim YW, Dessers E, Vrijhoef HJ. Instruments Measuring Integrated Care: A Systematic Review of Measurement Properties: Instruments Measuring Integrated Care. The Milbank Quarterly [Internet]. 2016 Dec [cited 2021 Mar 4];94(4):862–917. Available from: http://doi.wiley.com/10.1111/1468-0009.12233

53. Singer SJ, Kerrissey M, Friedberg M, Phillips R. A Comprehensive Theory of Integration. Med Care Res Rev [Internet]. 2020 Apr [cited 2021 Mar 4];77(2):196–207. Available from: http://journals.sagepub.com/doi/10.1177/1077558718767000

54. Wild E-M, Schreyögg J, Golubinski V, Ress V, Schmidt H. Hamburg Billstedt/Horn als Prototyp für eine Integrierte gesundheitliche Vollversorgung in deprivierten großstädtischen Regionen. Hamburg Universität; Gemeinsamer Bundesausschuss; 2020.

55. Larrain N, Wang S, Stargardt T, Groene O. Cooperation improvement in an integrated healthcare network: A social network analysis. Working Paper. 2022;

56. Rudawska I. Approaches to Evaluating Efficiency and Quality of Integrated Health Care: State of the Evidence. Ekonomia i Zarzadzanie [Internet]. 2016 Sep 1 [cited 2022 Jan 13];8(3):7–14. Available from: https://www.sciendo.com/article/10.1515/emj-2016-0019

57. Centers for Medicare & Medicaid Services C. Medicare ACO Shared-Savings Program Quality Measures [Internet]. No year [cited 2022 Jan 14]. Available from: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/ACO-Shared-Savings-Program-Quality-Measures.pdf

5. Appendices

5.1 Appendix 1. Reviewed Literature

First Author (year)	Source	Type of literature	Resources reviewed	Years covered
EU expert group (2017) [11]	Snowballing	Literature review + empirical research	12	N/A
Stiefel & Nolan (2012) [15]	Snowballing	Literature review + empirical research	>100	2007-2012
WHO (2016) [32]	Snowballing	Literature review + empirical research	19	N/A
WHO (2015) [26]	Snowballing	Empirical research	N/A	N/A
Raleigh (2014) [27]	Snowballing	Empirical research	N/A	
McDonald (2014) [8]	Snowballing	Empirical research	N/A	
Strandberg-Larse (2009) [48]	Snowballing	Literature Review	24	Until April 2008
Rudawska (2016) [56]	Literature Review on PubMed	Literature Review	13	1998-2003
Comendeiro-Maaløe (2019) [28]	Literature Review on PubMed	Empirical research	N/A	N/A
Levesque (2020) [20]	Literature Review on PubMed	Literature Review	19	Until 2017
Kaló (2020) [18]	Literature Review on PubMed	Case series	17	N/A
Pinter (2020) [21]	Literature Review on PubMed	Literature Review	87	2010-2018
Bevan (2019) [37]	HealthPros Network	Empirical research	N/A	N/A
Bos (2021) [30]	HealthPros Network	Empirical research	N/A	N/A
Wild (2021) [54]	HealthPros Network	Empirical research	N/A	N/A
Larrain (2021) [12]	HealthPros Network	Empirical research	N/A	N/A
Larrain (2022) [55]	HealthPros Network	Empirical research	N/A	N/A
Larrain (2022) [39]	HealthPros Network	Empirical research	N/A	N/A
Larrain (2022) [50]	HealthPros Network	Empirical research	N/A	N/A

Literature review on PubMed used the following Boolean search:
((["integrated"[Title/Abstract]) AND ("performance assessment"[Title/Abstract] OR "performance indicator*"[Title/Abstract] OR "performance evaluat*"[Title/Abstract]) AND ("healthcare"[Title/Abstract] OR "care"[Title/Abstract] OR "health"[Title/Abstract]))
Selection criteria: Performance assessment of integrated healthcare systems operating at system-level or describing performance assessment methods specific to integrated care.

5.2 Appendix 2. Quality criteria for the selection of performance indicators.

Criteria	Source
Importance and relevance	Raleigh et al. (2014) [30]
Validity	Raleigh et al. (2014) [30]
Accuracy	Raleigh et al. (2014) [30]
Reliability	Raleigh et al. (2014) [30]
Feasibility	Raleigh et al. (2014) [30] / Stiefel & Nolan (2012) [15]
Meaningfulness	Raleigh et al. (2014) [30]
Implications for action	Raleigh et al. (2014) [30]
Avoidance of perverse incentives.	Raleigh et al. (2014) [30]
Size of the population covered	Raleigh et al. (2014) [30]
Representation of important aspects of the care system	Raleigh et al. (2014) [30]
(Wholly or partly) within the control of care services i.e. attributability	Raleigh et al. (2014) [30]
Change detectable within suitable time frames	Raleigh et al. (2014) [30]
Unambiguous interpretation	Raleigh et al. (2014) [30]
Likelihood of being meaningful to users, carers and the public	Raleigh et al. (2014) [30]
Likelihood of being meaningful to care professionals, managers and commissioners	Raleigh et al. (2014) [30]
Reflecting the user perspective and/or value for money perspective	Raleigh et al. (2014) [30]
Timeliness	Raleigh et al. (2014) [30]
Ability to assess the impact on inequalities between user groups and areas in terms of access and ou	Raleigh et al. (2014) [30]
Measurable from routinely collected data.	Raleigh et al. (2014) [30]
Importance	Stiefel & Nolan (2012) [15]
Scientific acceptability	Stiefel & Nolan (2012) [15]
Usability	Stiefel & Nolan (2012) [15]
Are the measures actionable?	Bilheimer (2010) [31]
Are the measures sensitive to interventions?	Bilheimer (2010) [31]
Are the measures affected by population migration?	Bilheimer (2010) [31]
Are the measures easily understood by collaborating organizations, policy makers, and the public?	Bilheimer (2010) [31]
Is the meaning of an increase or decrease in a measure unambiguous?	Bilheimer (2010) [31]
Do the measures stand alone or are they aggregated into an index or summary measure?	Bilheimer (2010) [31]
Are the measures uniform across communities?	Bilheimer (2010) [31]
To what extent do measures address disparities as well as overall burden?	Bilheimer (2010) [31]
Can unintended consequences be tracked?	Bilheimer (2010) [31]
Simple, sensitive, robust, credible, impartial, actionable, and reflective of community values	Pestronk (2010) [30]
Valid and reliable, easily understood, and accepted by those using them and being measured by the	Pestronk (2010) [30]

Useful over time and for specific geographic, membership, or demographically defined populations	Pestronk (2010) [30]
Verifiable independently from the entity being measured	Pestronk (2010) [30]
Politically acceptable	Pestronk (2010) [30]
Sensitive to change in response to factors that may influence population health during the time that	Pestronk (2010) [30]
Sensitive to the level and distribution of health in a population	Pestronk (2010) [30]
Responsive to demands for evidence of population health improvement by measuring large sample sizes [57]	Pestronk (2010) [30]

5.3 Appendix 3. Compilation of performance indicators for integrated healthcare systems.

AVAILABLE AT: DOI 10.17605/OSF.IO/VNUZ6