CONNECARE

WP2 - CO-DESIGN OF INTEGRATED CARE

D2.3: PATIENT-BASED HEALTH RISK ASSESSMENT AND STRATIFICATION

H2020-EU.3.1: Personalised Connected Care for Complex Chronic Patients

Project No. 689802

Start date of project: 01-04-2016

Duration: 42 months

Project funded by the European Commission, call H2020 – PHC – 2015	
☑ **PU**	**Public**
☐ PP	Restricted to other programme participants (including the Commission Services)
☐ RE	Restricted to a group specified by the consortium (including the Commission Services)
☐ CO	Confidential, only for members of the consortium (including the Commission Services)

Revision: 25

Date: 28/07/2017

Document Information

Project Number	689802	**Acronym**	CONNECARE
Full title	Personalised Connected Care for Complex Chronic Patients		
Project URL	**http://www.CONNECARE.eu**		
Project officer	Hubert Schier		

Deliverable	**Number**	2.3	**Title**	Patient-based Health Risk Assessment and Stratification
Work Package	**Number**	2	**Title**	Co-design of Integrated Care

Date of delivery	**Contractual**	MONTH 16	**Actual**	MONTH 16
Nature	Prototype ☐ Report ☑ Dissemination ☐ Other ☐			
Dissemination Level	Public ☑ Consortium ☐			

Responsible Authors	Josep Roca & Akos Tenyi	**Email**	JROCA@clinic.cat
Partner	IDIBAPS	**Phone**	+34 93 227 5747
Contributors	Rachelle Kaye (ASSUTA); Maarten Lahr (UMCG); Stefano Mariani (UNIMORE); Isaac Cano (IDIBAPS); Jordi de Batlle (IRBLL); and, Gerard Torres (IRBLL)		

Abstract	This CONNECARE document (D2.3) has a threefold aim. Firstly, to describe the consensus achieved by the clinical partners of the consortium regarding conceptual and pragmatic aspects of health risk assessment. The document proposes an operational formulation of enhanced clinical risk predictive modelling to be adopted in CONNECARE (Task 2.3). It has been worked out iteratively with Task 3.4 for elaboration of clinical decision support systems (CDSS) supporting dynamic risk assessment using multilevel data sources. Section 2 of the document describes the maturity of each site regarding risk stratification. For practical purposes Lleida and Barcelona have been collapsed as one regional development (Catalonia, ES). Finally, Section 3 indicates the overall CONNECARE outcomes regarding Task 2.3, as well as the need for specific PDSA approaches to health risk assessment in each site. The document defines the basis of the new predictive modelling approach that will be assessed in the three case studies deployed in CONNECARE (WP6).

Table of contents

Executive Summary

Health risk assessment is a relevant component in the strategies for regional adoption of integrated care because of its impact on the design of healthcare services *(**population-based** health risk assessment)*, as well as for enhanced clinical management *(**patient-based** health risk assessment)*.

Patient management purely based on **clinical criteria** *(professional training, knowledge, instinct and experience)* or combined with **rules-based clinical management** *(thresholds for certain parameters defining pre-established decision criteria)* constitute most of current health professional practice. In contrast, the use, on a regular basis, of **predictive modelling tools** for clinical decision support *(predictive modelling establishing relationships between sets of variables and outcomes generated using statistical or machine learning tools)* is still in its infancy despite it seems a natural step towards customization of care to patient's needs.

The document focuses on the description of CONNECARE strategies to optimize patient-based health risk assessment and service selection in order to facilitate elaboration of individual care plans in the clinical scenario. During the project lifespan, two use cases will be addressed: i) **complex chronic patients** (CONNECARE case study 1); and, ii) **peri-surgical care** (CONNECARE case studies 2 & 3). One of the core hypotheses to be explored in these use cases is if information contained in population-health risk assessment may significantly contribute to enhance clinical risk prediction, as formulated in [1].

Section 1 of the deliverable describes the current state of the art of both population-based and patient-based health risk assessment, as well as their desirable articulation to facilitate clinical decision support for optimal service selection. Moreover, the section analyses specific proposals for prompting evolution of current healthcare towards the concept of **learning healthcare systems** [2].

Section 2 describes existing practices regarding health risk assessment in the four CONNECARE sites. For practical purposes, Lleida and Barcelona are described as one area because the two sites share a common regional policy in terms of risk assessment. The section also indicates future plans and expected benefits for implementation of predictive modelling for clinical decision support at site level, as well as main barriers and/or facilitators identified for such implementation.

Finally, **Section 3** formulates a common CONNECARE conceptual frame for health risk assessment and service selection, recommends methods for evaluation and explores potential implementation strategies at site level.

The information included in the current document covers the aims of Task 2.3 *(Health risk assessment for patient stratification)*. It has been elaborated through repeated iterations with the team responsible for the design of CONNECARE Adaptive Case Management *(WP2 – D2.2.Adaptive Case Management Design)* and Clinical Decision Support Systems (CDSS) *(WP3 – D 3.2: First screening and risk stratification DSS)*. These deliverables will enrich the clinical studies (WP6) and should provide key elements to generate final recommendations for patient-based health risk assessment and service selection in *WP7 – D7.4: Recommendations of final services and proposals for scale-up integrated care.* This will contribute to the general goal of the project - which is: to foster large scale deployment and adoption of integrated care services through significant contributions to healthcare value generation and by facilitating reductions of outcome variability within and among European regions.

1. Health risk assessment, stratification and service selection

Decision making in the clinical setting traditionally faces two types of challenges, with the individual patient as the recipient of action. On the one hand, reactive care aims to solve specific disease events, such as pneumonia, appendicitis, acute episodes on the backdrop of chronic diseases (i.e., chronic obstructive pulmonary disease (COPD), cardiac failure, type I or type II diabetes mellitus,…), etc…. On the other hand, the design of mid- and/or long-term therapeutic plans for a given patient also constitutes a core medical activity.

Clinical decisions are traditionally based on: i) knowledge, general and field-specific, of the health professional, ii) previous experience, as well as, iii) intuition. It is of note that rule-based decision making relying on robust medical evidence, often generated through randomized controlled trials, is already an essential component of the professional knowledge contributing to efficiency of the decision making process.

Since early 2000s, two key phenomena are prompting substantial changes of both clinical research and practice. Firstly, systems biology methodologies (i.e. holistic approach, use of computational modelling, etc…) are being progressively embedded in medical practice shaping the practicalities of systems medicine [3–7]. Simultaneously, the use of information and communication technologies (ICT) is facilitating access to an enormous amount of patient-related information, as well as providing sophisticated computational and machine learning tools that offer high potential for predictive modelling [8–18]. These facts should increase the potential for assisting health professionals in patient-based (or clinical) decision making.

It is of note that current strategic changes in the orientation of clinical practice are fully aligned with the vision of systems medicine. That is, a **holistic understanding of disease-related events**, as well as a **preventive approach to medical practice,** which necessarily involves a multidisciplinary team-based orientation of therapeutic strategies with a distribution of roles and tasks among professionals often working in different healthcare tiers. This is a key component of the CONNECARE Adaptive Case Management design *(WP2 – D2.2.Adaptive Case Management Design)*. The new scenario is triggering emerging computational requirements in terms of predictive modelling for individual-based decision making, moving toward predictive and personalized medicine. The above changes are heavily accelerated by well identified XXI-century factors such as: high prevalence of chronic conditions, ageing and need for cost-containment, aiming at financial sustainability of health systems. All in all, it

seems clear that adoption of the new health paradigm is requiring convergence between practicalities associated with large scale implementation of integrated care, based on the Chronic Care model [19], and stepwise deployment of a system medicine approach in clinical practice.

In accordance with the holistic approach alluded above is the consideration of the need for a population-based analysis of health-related events; that is, predictive modelling at population level. Under the novel conceptual frame, population-based health risk assessment should not only be considered as a tool for healthcare service design/commissioning and/or financial optimization. Instead, we propose to consider it as complementary to patient-based health risk assessment. In this regard, the current document (**Section 1.1**) heavily endorses the exploration of potential synergies between population-based and clinically-oriented health risk prediction.

The document also examines the practicalities of the relationships between health risk assessment and service selection in clinical practice (as described in **Section 1.2** and illustrated in **Figure 1**). In other words, we are taking a highly pragmatic approach to health risk prediction. We envisage stratification, either population-based or clinically-based, not only as a conceptual issue for classificatory purposes; but as a way to optimize health of both citizens living in a given geographical area (**population-health** approach) and individual patients in the clinical scenario (**patient-based** approach).

Finally, we are interested on transformational aspects of current health systems fostering a systems medicine approach to integrated care that can foster sustainable productive interactions between clinical practice and generation of knowledge. Recent recommendations for implementation of Learning Health Systems (LHS), as those generated by the American Heart Association (AHA) [2], are analysed in **Section 1.3** as promising initiatives to pave the way for adoption of the new health paradigm.

Overall, the document is highly aligned with Task 2.3 of the CONNECARE project addressing health risk prediction by combining an already operational population-health predictive modelling tool elaborated in Catalonia (Adjusted Morbidity Groups, GMA) [1,20–22] and the patient-based scoring strategy generated within the project. This task aims to set-up enhanced clinical predictive modelling to be assessed in the three case studies deployed in WP6.

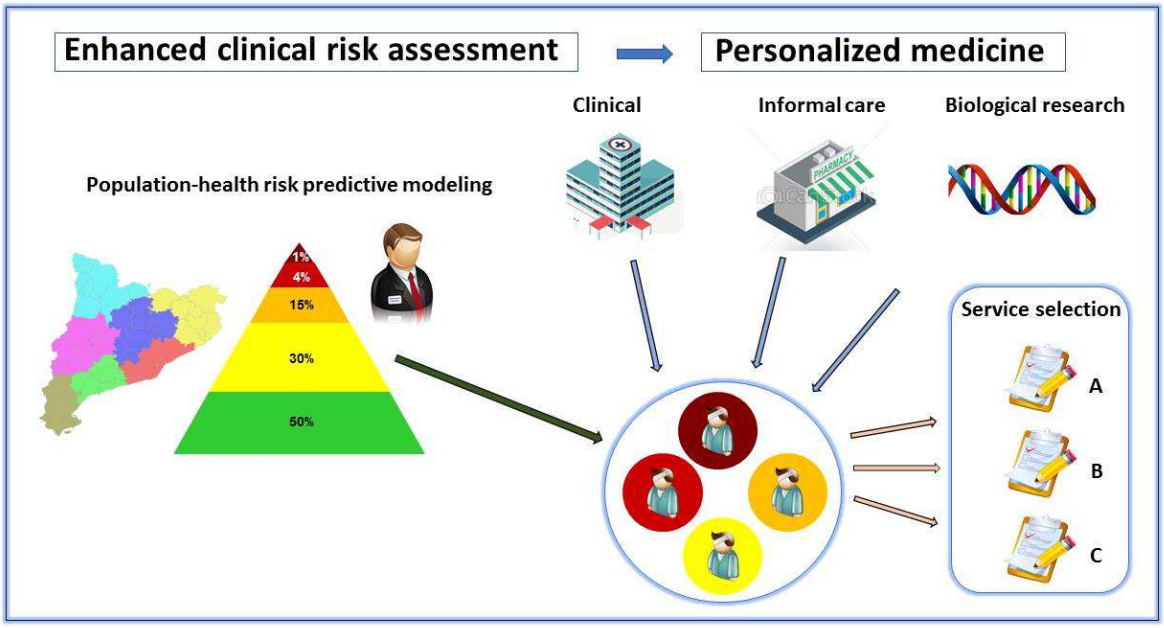


Figure 1 – In the document, **patient-based health risk assessment** is used as synonymous with **enhanced clinical risk assessment**, which adopts of a holistic approach that fosters inclusion of covariates from multilevel data sources, namely: i) Clinical, ii) Informal care; iii) Biological research; and, iv) Outcomes from population-health risk predictive modelling, resulting in enhanced patient-based stratification and optimization of service selection. This approach paves the way towards personalized medicine. **Population-health risk predictive modelling** includes all the citizens in a given geographical area. Left-hand-side of the figure displays the Catalan risk stratification-pyramid in 2014. It should be distinguished from **population-medicine predictive modelling** which is elaborated from groups of patients with pre-defined inclusion/exclusion criteria. We hypothesize potential synergies between population-health predictive modelling and clinical risk assessment, as explained in the document.

1.1 Health risk assessment

Patient-based health risk assessment - In the clinical management domain, risk prediction of well-defined medical problems aims to support health professionals in the decision making process for a given patient. The definition is valid for the process solving a clinical episode of pneumonia or exacerbation of a chronic disease, for example, but also to define mid- or long-term action plans for a chronic patients aiming at developing an optimal care plan.

However, the novel healthcare scenario is prompting action toward enhancement of clinical health risk predictive modelling using currently available multiple sources of information; by,

adopting a multilevel approach as displayed in **Figure 2**, in order to anticipate, for the individual patient, the development or progression of disease to prevent, or reduce its impact: i) through early diagnosis; and, ii) triggering cost-effective preventive interventions.

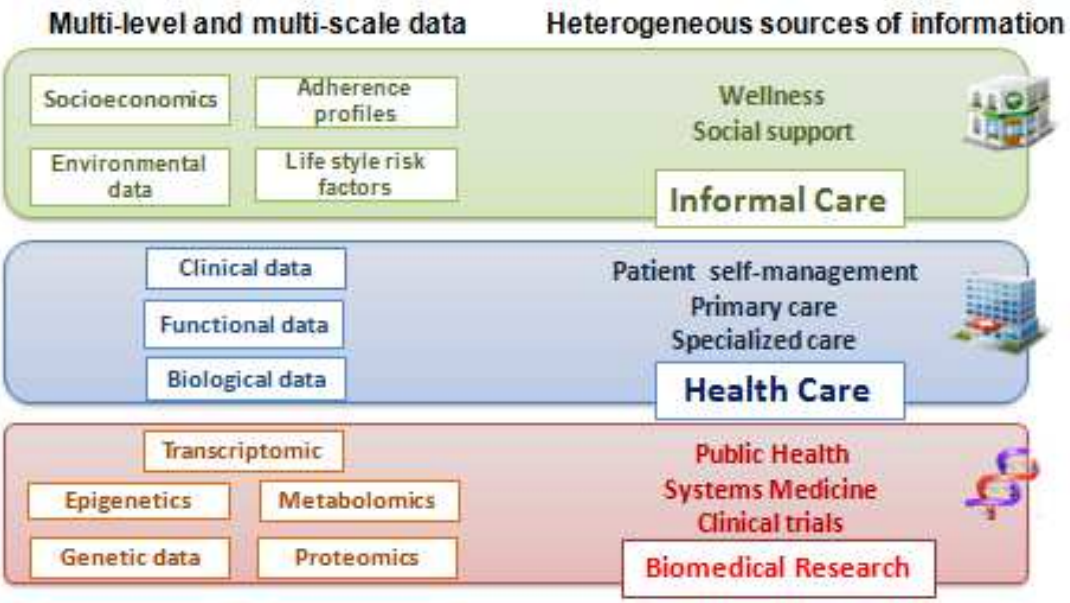


Figure 2 – The dimensions of patient health indicated in the figure may contribute to the enrichment of clinical risk predictive modelling. As a first step, we propose including the outcome of the population-health risk assessment as a covariate in clinical risk predictive modelling. For future personalized care for chronic patients, enhanced dynamic communication among Informal Care, Health Care and Biomedical Research will allow inclusion of several dimensions into clinical risk predictive modelling. It will be done through multilevel/multi-scale heterogeneous data integration into a Digital Health Framework, as depicted in **Figure 3**. This figure was adopted from [8].

Current status and factors limiting evolution – While adoption of rule-based clinical decision support has made substantial progresses over the last years, the use of computational models for patient-based health risk predictive modelling to enhance clinical decision support is still in its infancy. We identify three main factors to be considered in order to overcome current limitations of patient-based risk prediction. Firstly, enrichment of predictive modelling based on clinical covariates with other variables (i.e. informal care, population-health stratification, biological data, etc…) obtained with a multilevel approach **(Figure 2)**. A second factor is to ensure general applicability and transferability of current predictive tools, addressing specific clinical issues with a high predictive power, to other populations outside the source study

groups. Last, but not least, there is a clear need to overcome limitations of use of risk factors as prognostic factors, as described in [23]. The analysis of how to make these three limiting factors operational will be addressed in **Section 1.2** of the current document.

Population-based health risk assessment – Risk assessment in the health services domain, differentiated from individual clinical risk prediction, targets groups of citizens or patients. It provides tools for policy makers and/or payers that are very useful for service commissioning or other uses like risk adjustment analyses or actuarial approaches. The last two items are beyond the focus of the current document.

It is of note that a clear distinction among **population-based risk assessment tools** should be done depending upon the characteristics of the source population. On the one hand, health risk assessment tools generated by modelling the entire population of a given region (or geographical area) with a holistic approach are considered to follow a **population health** approach, as proposed by Kindig D et al. in 2003 [24,25]. A population-health approach facilitates elaboration of stratification maps characterizing risk strata distribution of the entire population in a given geographic location. It allows identification of subsets of citizens with similar healthcare requirements facilitating both case finding and screening. The former, case finding, identifies highly vulnerable patients, allocated at the tip of the risk pyramid who are prone to major deleterious health events such as unplanned hospital admissions/re-admissions, fast functional decline and/or death [26,27]. Likewise, performing screening for discovery of cases with non-manifest illnesses may benefit from early diagnosis and cost-effective preventive interventions [28]. On the other hand, health risk assessment derived from subpopulations of patients are regarded as following a **population medicine** approach [24]. Comprehensive descriptions of the characteristics of health risk predictive modelling and the logistics required for deployment are reported elsewhere [29–33].

Current status and factors limiting evolution – A comprehensive analysis of the characteristics of the relevant population-based risk prediction tools was carried out by two EU projects: ASSHES [34,35] and ACT [1,36]. Briefly, population-based health risk predictive modelling tools in place in leading European regions display marked heterogeneities in several dimensions that preclude comparability of the risk pyramid distributions across regions. Moreover, current population-based tools show some intrinsic limitations in terms of robustness of derived estimations, not precluding their usefulness for policy making.

It is of note that transferability across regions, and potential for flexibility, are identified as two key requirements for any population-based health risk assessment tool. Factors such as: (i)

license binding constraints, (ii) proprietary software; (iii) lack of availability for inspection; and/or, (iv) rigidity of some computational algorithms (i.e. due to inclusion of expert-based criteria in some morbidity groupers) are limiting transferability for most of the existing population-based risk assessment tools.

The above factors are also precluding adaptation of the current risk prediction tools toward evolving requirements such as: (i) integration between healthcare and social services; and, (ii) implementation of synergies between population-health and clinically oriented risk predictive modelling, as described in [1].

Toward population-health approaches – There is a general consensus on the existing intrinsic limitations of population-based risk prediction based on a population-medicine approach. Also, there is agreement on the positive role ascribed to population-health risk assessment in regional deployment and adoption of integrated care services. We observe that well-known classical population-medicine risk prediction tools, like SPARRA [37], are quickly evolving toward a population-health approach.

In this regard, we suggest two basic pillars for a future European adoption of population-based risk prediction: i) implementation of the recommendations for risk predictive modelling tools displayed in **Table 1**; and, ii) ability to report on the list of basic indicators depicted in **Table 1**. The current heterogeneities among regions clearly indicate that adjustment of the current settings to the recommended good practice will require site-specific transitional strategies whose common goals and basic principles are indicated in **Tables 1** and **2**. Key operational steps needed for practical implementation of a regional strategy for population-health risk predictive modelling are summarized in **Table 2**. Irrespective of the management modalities associated with generation and exploitation of population-health risk predictive modelling (public vs private funding, supply & management), we emphasize the need for openness, flexibility and transferability of population-based risk predictive modelling in order to fulfil their core purposes. However, we acknowledge the complexities of the issue, also involving legal and ethical aspects requiring proper regulation, irrespective of the finally adopted business model for commercial exploitation of the predictive modelling tools.

Table 1 Recommendations for good practice population health risk assessment [37]

Domain	Recommendations	Level of evidence
Type of risk stratification tool	Predictive model using a population health approach	High [24,25,29–32,38]
Validation of the model	Longitudinal follow-up	High [23]
Predicted/explained Outcomes	Unplanned hospital-related events; risk of institutionalization; Death; case prognosis	High [24,25,29–32,38]
Source sample	Whole regional population	High [24,25]
Statistical model	Predictive modelling	High [24,25,29–32,38]
Statistical indices	Standardization on reporting performance (positive predictive value, PPV)[23] and sensitivity across risk bands	Moderate [23] (*)
Population usefulness	Risk adjustment; planning and commissioning health services Support to novel reimbursement models	High [33,39,40]
Clinical & social usefulness	Identification patients at high risk and cost-effective preventive clinical & social interventions	High [24,25,29–32,38]
Periodicity of updates	Semester	Low (**)
Clinical accessibility	Available into the professional workstation through clinical decision support systems	High (***)
Flexibility & Transferability	Open algorithms, open source, reduced or no licence binding. Morbidity grouper based on statistical criteria adjusted to the target population.	High

(*) *To report metrics indicating sensitivity/specificity of predictions is recommended for good practice. But, some regions adopt a pragmatic approach classifying individuals into the risk-strata pyramid without sensitivity/specificity because of rather poor robustness of predictions provided by most of the models.*

(**) *Periodicity of updates depends on the logistics available in each site. Yearly or six-month basis seem reasonable*

(***) *Development of adequate clinical decision support systems (CDSS) depends on three main factors: i) Robustness of computational modelling feeding the CDSS; ii) Refinement of the CDSS generated by the clinical feedback; and, iii) Appropriate dashboard providing a user-friendly interface*

Table 2. Recommended operational steps toward implementation of a regional strategy for health risk assessment [1].

Recommended operational steps
1. **Health risk predictive modelling implementation** - Use a population-health risk assessment tool fulfilling the requirements indicated in **Table 1**, either by fostering the evolution of your own risk assessment tool or by adopting an existing risk assessment tool that fits your local needs, which can be used without any license bindings and supports an open market of suppliers. Screen your population on a regular and repeated basis. Be aware of the logistics required at regional level to develop operational health risk prediction strategies: i) identify and overcome the practical local hurdles and barriers for accessing and linking routine administrative and clinical data and, ii) estimate the cost of running a tool, software platform, data integration, as well as labour for operations.
2. **Define and activate specific functionalities -** Use population-health risk stratification to understand the needs and risks of your population to target and prioritize effective integrated care. Make the outcome to be predicted operational (risk type: unplanned hospital related event; functional decline/frailty; death, etc...) aiming at healthcare value generation by embedding risk assessment into healthcare delivery (i.e. setting cost-effective preventive interventions). Also, decide what risk strata you would like to address (i.e., risk pyramid with one top, two intermediate and one bottom layer).
3. **Engage professionals and customize the setting** - Engage and educate your healthcare professionals and clinical staff in the use, value and shortcomings of risk stratification in order to gradually obtain the buy-in of the clinical community. Use an iterative co-design process involving healthcare professionals to define clinical applicability of outcomes of population-based risk prediction. Also, involve them in designing the characteristics of the dashboard displaying information on risk outcomes in the clinical workstation. Likewise, cohorts and associated protocols designed to assess interventions on specific risk strata should be implemented in close collaboration with healthcare professionals who should be informed about usefulness and potential pitfalls associated with health risk prediction. Moreover, studies evaluating the potential of population-based risk assessment for enriching individual risk predictive models addressing specific clinical issues should be designed and conducted with clinical professionals.
4. **Generate recommended indicators with standardized reporting** - Protocols for data harmonization and data reporting should be in place and shared at European level in order to ensure comparability across regions.

A comprehensive approach for health risk assessment

Synergies between patient-based and population-health risk assessment – As alluded to above, these two modalities of health risk assessment must not be addressed as isolated silos. There is increasing evidence [41] indicating that those population-health risk predictive tools following the principles indicated in **Table 1**; thus, showing flexibility and transferability, could be further statistically refined by enriching the spectrum of covariates considered in the patient-based health risk assessment computational models.

An example of this would be personalizing information on socioeconomic status or by taking into account other relevant additional variables. Moreover, as an effective bridge between clinical risk assessment and population-health predictive modelling, we propose to incorporate the classification of the individual in the population-health risk stratification pyramid as one of the covariates of patient-based health risk predictive models. It is of note, however, that prospective assessment of the practicalities, as well as quantification of the added value of the proposed approach, needs to be further examined with targeted research strategies.

Paving the way for personalized medicine – As alluded to above, current patient-based health risk predictive models are essentially using clinical variables only. However, three categories of covariates have been identified to show potential for inclusion into patient-based health risk predictive models, as displayed in **Figure 1**: (i) input from enhanced case finding tools; that is, population-health risk predictive models, as mentioned above; (ii) individual clinical, physiological and biological information relevant to the medical problem being assessed; and (iii) subject-specific informal care data including lifestyle, adherence profile, socioeconomic status, requirements in terms of social support and environmental factors. It is hypothesized that inclusion of all these covariates influencing patient health may markedly increase the predictive accuracy and facilitate clinical decision-making based on sound estimates of the prognosis of an individual.

The three categories of covariates displayed in **Figure 2** shall be dynamically captured from different sources, respectively: (i) population-health risk predictive models; (ii) articulated healthcare and biomedical research knowledge (integration of clinical, physiological and biological/molecular information); and, (iii) in-place personal health folders (lifestyle, adherence profile, socioeconomic status, social support and environmental factors).

The implementation of specific solutions within a Digital Health Framework, conceptually formulated in [8] (**Figure 3**), should have the potential to articulate the three categories of

variables potentially allowing for dynamic assessment of health risk both for population-based purposes, but also for specific clinical problems. A Digital Health Framework is only a conceptual formulation, but it contains the seeds to foster the concept of the "exposome", as defined by Coughlin SS [42], which provides the basis for personalized medicine for chronic cases. There is no doubt that the implementation of specific solutions within the envisaged Digital Health Framework constitutes an ambitious endeavour requiring a stepwise approach to effectively overcome major challenges involved in the transitional process to make it operational.

Enhanced applicability and integration of powerful data analytics, including risk predictive modelling, into clinical practice also constitutes a central goal of the CONNECARE project. In this regard, the development of novel clinical decision support systems, supported by advanced visual analytics, facilitating representation of patient information for effective clinical management of time-varying individualized data is a real yet unmet need to facilitate clinical judgement for decision-making. Moreover, studies assessing the potential of different modalities of patient gateways, like the personal health folder, for patient self-management purposes and for collection of informal care variables, are urgently needed.

Current status and factors limiting evolution - This holistic approach generates novel requirements to be adopted by the field. Firstly, the need for multilevel integration of heterogeneous patient information, namely: socio-economical, life-style, behavioural, clinical, physiological, cellular and "omics" data [8,43], and their use for the study of disease mechanisms. Secondly, the need to extend current trends on open data from the biomedical community [43] to the clinical practice and the whole society, by engaging citizens and solving privacy and regulatory constraints.

The novel healthcare scenario reveals new emerging needs regarding highly relevant non-solved ethical issues. These are related to privacy, security of data transfer, as well as risks associated with healthcare decisions that rely on inadequate risk predictive models. The complexities involved in some of these aspects can only be addressed through a democratic debate; openness and transparency of the healthcare governance; as well as a timely and appropriate evolution of legal frames.

Main identified barriers and opportunities to enable the required potential for Big Data analytics in health applications in health have been identified in a European Union study on

Big Data in Public Health, Telemedicine and Healthcare [44]. As a result of the systematic review recommendations were identified ten relevant fields, which should be taken into account to structure the following list of potential areas of improvement, as recently reported in [43] and summarized in **Table 3**.

Table 3. Barriers and opportunities to enable Big Data applications in health [43].

Barriers and opportunities
1. Standards and protocols - Health data is not always available in a digitized form. Its transformation into structured formats (e.g., HL7 CDA) and the movement of health registries out of current silos in formal care, informal care and biomedical research might be costly. Moreover, current developments focus on standards to guarantee data standardization and interoperability (e.g., ICD, DICOM, SNOMED, HL7, ISATab, etc.), but do not consider data quality and how to manage patient identity across data sources (e.g., unique patient identifiers)
2. Technological developments - New technological and software developments can improve the utility and security of health registries and enable data analysis in real-time settings. However, in order to run pre-processing routines and machine learning algorithms to build predictive models and perform integrative multi-scale simulations, it arises the need to allocate clusters of computers working in a collaborative way and supporting novel stacks of privacy-preserving software frameworks and tools which require expert Big Data scientists and engineers.
3. Data analytics - High awareness and understanding of the added-value of Big Data applications with Health information can promote the development of success stories. Considering that real-time, menu-driven, user-friendly and transparent data analytics tools might not be fully developed yet, entrepreneurs and early- adopters might foster the use of innovative Big Data analytics in health.
4. Legal aspects - Although the General Data Protection Regulation (EU) 2016/679 provides more precise definitions of health data, consent and scientific research, most rules relevant for health (such as the eventual requirement of informed consent, the potential use of professional secrecy as an obstacle to share health information, and the many references to member state laws) might hinder gathering and sharing personal health data. Therefore, there is an urgent EU need for aligning existing fragmented national legislations on collection, storage, analysis, use and dissemination of health data toward the foundation of global legal frameworks to support development and assessment of digital health services.
5. Stakeholders - There is an increasing need for the coordination of interests and responsibilities among different stakeholders (e.g., payers, healthcare providers, academia, clinicians, patients and patients' associations, etc.). Involving opinion leaders in different public and private stakeholders' groups in public consultations might reduce risk while increasing acceptance and the probability of successful applications.
6. Business models - Huge potential health and economic benefits can be envisaged in terms of accelerating cross-fertilization between knowledge generation (biomedical research) and both health and informal care data. Progress in this direction will be strongly associated to innovative business models, such as bundle payment for care improvement, providing sustainability of platforms beyond specific projects that triggered the initial settings.

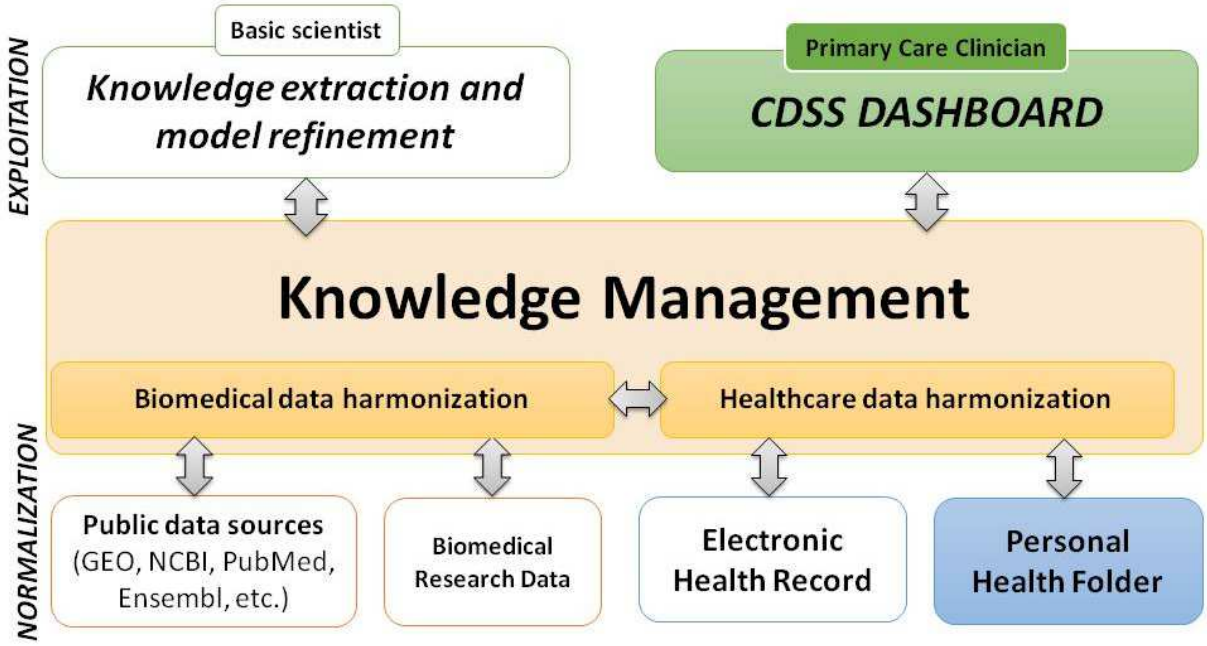


Figure 3 – Scheme of the Digital Health Framework [8], a digital data normalization and knowledge management framework for knowledge generation and to embed novel Clinical Decision Support Systems (CDSS) into integrated care processes.

1.2 Patient-based risk stratification, patient similarity and service selection

Section 1.1 provides the conceptual basis to overcome, at least partly, the three current limitations of patient-based risk prediction alluded above. It should be done through: i) Enrichment of predictive modelling with other variables obtained with a multilevel approach; ii) Ensuring general applicability and transferability of current predictive tools to other populations outside the source study groups; and, iii) Prevent the use of risk factors, defined as odds ratios, to assess prognosis in clinical practice. Instead, the receiver-operating characteristic (ROC) curve should be the standard tool for assessing performance of clinical risk prediction models [23]. This section addresses advanced operational aspects for a future generation of personalized patient actions plans.

Rationale – When considering one disease in a given patient, prognosis is essentially based on two main parameters: i) **severity**, defined as the degree of alteration of the organ/systems caused by the disease, which have an impact on the functional reserve; and, ii) **activity**, defined by the rate of progression of the disease. It is of note that appropriate markers of these two phenomena contribute to define both risk and prognosis of the patient which, in

turn, facilitates his/her classification into risk strata; that is, patient stratification. Moreover, the combination of patient stratification and the identification of the disease end-points, or target outcomes, are the two key elements to define specific therapeutic strategies or action plans for the patient. The ultimate aim is to classify the patient in the appropriate health tier, and identify the type of service that would allow optimization of healthcare provision.

However, real world healthcare settings face high levels of complexity that are imposing huge challenges, specifically on risk assessment and patient stratification for adequate service selection. Main determinants of such complexity are: i) Patient heterogeneity with lack appropriate biomarkers and/or insufficiently defined end-points of the disease; ii) Co-existence of one main disease and several accompanying disorders (or co-morbidities) or simply multi-morbidity; in some cases showing shared mechanisms that may explain co-morbidity clustering [45]; iii) Poor control on factors determining health status beyond the clinical scenario (socio-demographics, biological and lifestyle related data, as displayed in **Figure 2**); iv) Patient health risk is a dynamic phenomenon with sometimes unexpected events that requires high levels of flexibility in terms of event-handling by the case manager in charge of the patient; and, v) Fragmentation of healthcare services.

CONNECARE assumes that convergence between large scale implementation of integrated care and adoption of a systems medicine approach should contribute to successfully make management aware of the above healthcare complexity, as represented by complex chronic patients. It is of note, however, that both dynamic health risk assessment and innovative organizational settings that facilitate collaborative adaptive case management (ACM) need to be implemented, as described in *D2.2. Adaptive Case Management Design*.

Regarding the methodological aspects of patient-based risk assessment, recent recommended virtual cohorts based on patient similarity [46,47], seem to provide innovative approaches to stratification of patients leading to more accurate risk prediction and better service selection than that achieved with classical tools, as described below.

Methods and tools - Traditional predictive modelling uses routine data which are entered into a statistical model, mostly into a logistic regression model, in order to calculate the risk through odds ratios, e.g. risk of hospital readmission. However, implementation of a predictive model requires a definition of the risk threshold score identified using ROC curves [23]. This is the level above which people are to be defined as 'high risk' and above which services or interventions are to be put in place. The accuracy of prediction of healthcare utilization and

costs (i.e. ER and hospital admissions, and high-cost users) of currently available models ranges from 0.68–0.81 (area under receiver operating characteristic [ROC] curve) [23,48,49].

However, many advances have been made in predictive modelling toward outcome prediction. These innovations target an average patient and are insufficiently adjustable for individual patients. One developing idea in this field is individualized predictive analytics based on patient similarity. Patient similarity is a central concept for quantifying the degree of similarity between an index patient and a past patient regarding their risk factors, including biological and clinical characteristics. This approach aims to identify similar patients from the historical data and derive insights from their records to provide personalized predictions [46,47,50]. Virtual patient cohorts can be created using clustering methods that find groups of similar patients in the population or personalized cohorts that can be assigned to an index patient by retrieving the N most similar patients. Stratification of patients based on several factors including diagnoses, risk factors and biological background could allow optimization of healthcare provision and could address main limiting factors arising from the high complexity of the task. Examples of current applications of this approach include prediction of heart failure from tele-monitoring data [51,50], risk factor identification of similar patients [52] and personalized treatment and drug recommendation systems [50,53], which list could be further broadened in the DHF scenario.

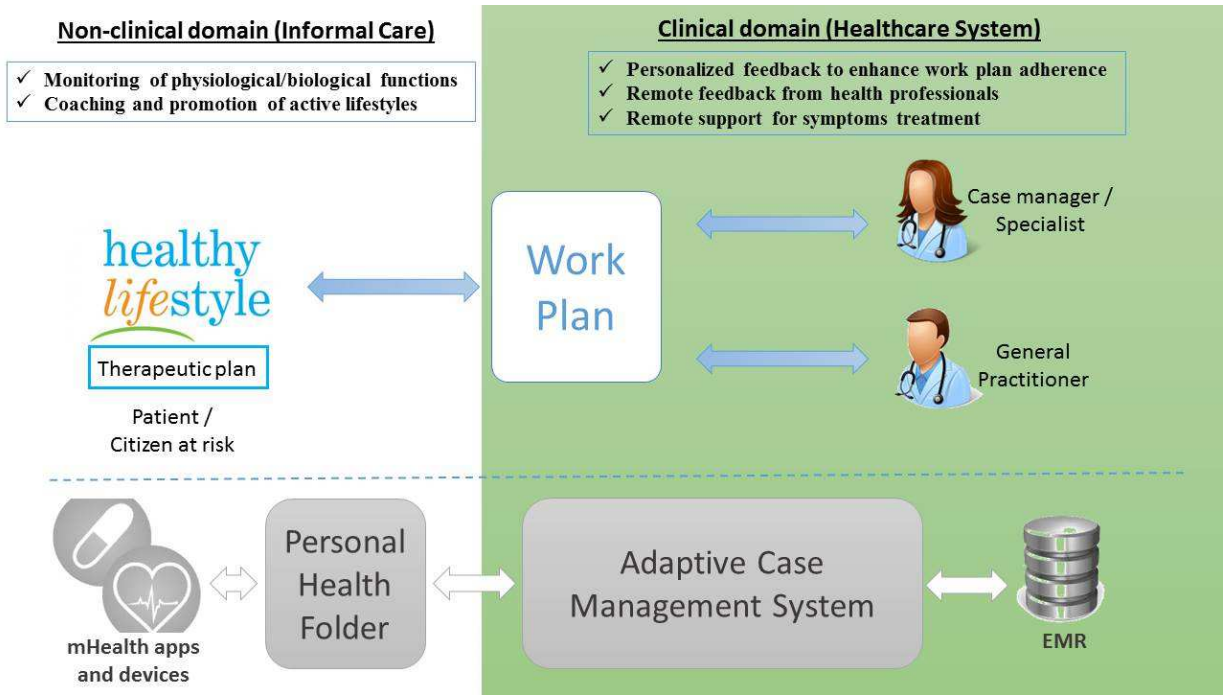


Figure 4. The figure shows two interoperable domains (Informal Care and Healthcare) with technological elements providing support to execution of the action plan including: i) Patient self-management; ii) Promotion of healthier lifestyles; and, iii) Remote interactions with health professionals. On the left, there is Informal Care area with the patient having access to the Personal Health Folder (PHF) wherein she/he can answer questionnaires, perform monitoring through mHealth apps, and have access to a follow-up reports and tailored educational information, as defined in the work plan (centre of the figure). On the right, we see the Formal Care domain wherein the case manager and/or other health professionals have access to an adaptive case management system for work plan prescription, follow-up and coaching. The adaptive case management system supports execution of the patient work plan and provides a bridge of interoperability and collaborative tools among the patient (through the PHF), the case manager and the electronic medical record (EMR).

Comprehensive & dynamic individualized care plans – In an integrated care scenario, like CONNECARE, individualized care plans should address multi-morbidity; that is, they should be patient-centred (not disease-oriented). Dynamic health risk assessment properly implemented and promotion of patient-self management aiming at optimal adherence should be a well-defined end point. Therapeutic plans should be designed with a preventive approach while encompassing pharmacological and non-pharmacological actions. Moreover, accessibility to health professionals, as well as to informal careers, with ability to perform off-line, and sometimes on-line, remote interactions, including monitoring, should be ensured. We believe that a proper interplay between the personal health folder (PHF) (including the self-

management system - SMS - developed in CONNECARE) and the healthcare system may facilitate the ideal setting described above and illustrated in **Figure 4**. Moreover, the PHF can be an optimal interface to integrated informal care information into the healthcare scenario, as envisaged in the concept of Digital Health Framework (DHF).

1.3 Learning Healthcare Systems

We acknowledge that implementation of the setting described above implies realization of a new health paradigm. Some of these elements, already targeted in the CONNECARE project, constitute key achievements toward the concept of learning healthcare systems recently described by the American Heart Association (AHA) [2].

In 2013, the Institute of Medicine reported Best Care and Lower Cost: The Path to Continuously Learning Health Care in America [54] wherein the concept of Learning Health Care Systems (LHS) was formulated as a strategy to improve the quality and efficiency of healthcare. A recent document generated by the AHA [2] further develops the concept of LHS and proposes specific steps to make it operational and evaluate its implementation, see Table 2 in [2]. It is of note that LHS is fully aligned with the concept of DHF displayed in **Figure 3** of the current document.

Briefly, LHS uses health information technology and the health data infrastructure to apply scientific evidence at the point of clinical care while simultaneously collecting insights from that care to promote innovation in optimal healthcare delivery and to fuel new scientific discovery [2]. Such a system creates an iterative learning process where evidence informs practice and practice informs evidence (**Figure 5**).

The main goal of LHS is to facilitate an optimal care decision and delivery by reducing the complexity of the massive amount of clinical data that's being produced every day and to improve efficiency of health outcomes both in terms of well-being and expenditures. The LHS relies on the availability of health-related data and tools that process it, such as predictive modelling and clinical decision support (CDSS) contributing to the acceleration of evidence diffusion to practice, help to identify gaps in care and to target interventions to appropriate population.

Main technical building blocks of such a system are data availability, predictive modelling, service selection and clinical decision support systems. Data is key because it provides the continuous feedback from the practice, while predictive modelling processes and extracts

important information from the produced data. Computed patient risk then can be used to stratify patients to intervention groups that help in the optimal service selection for the patient.

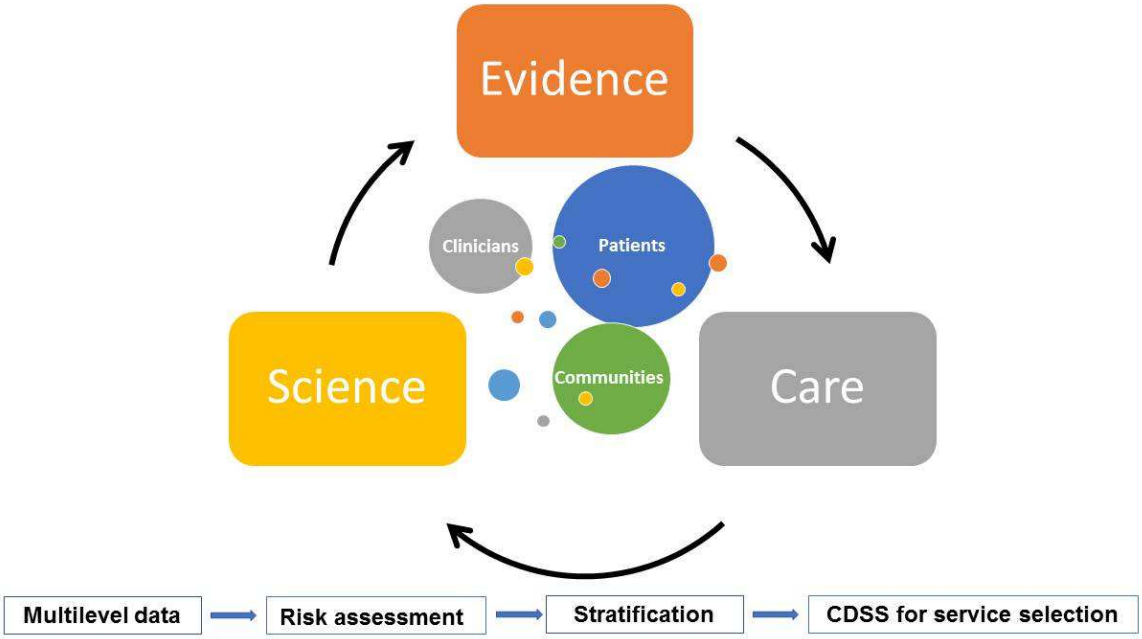


Figure 5 – **Basic traits of a Learning Healthcare System (LHS)**. It constitutes an organizational concept technologically supported by the Digital Health Framework (DHF) depicted in **Figure 3** of the current document. The LHS fosters generation of scientifically-based evidence and speed-up its applicability into healthcare. Regarding health risk predictive modelling, it shall allow inclusion of covariates from multilevel data sources which should enhance model robustness and eventually transferability feeding clinical decision support systems (CDSS) for appropriate integrated care service selection.

2. The sites: current status and implementation plans

2.1 Catalonia (ES): Lleida and Barcelona

Catalonia is actively working to achieve maturity of ongoing developments aiming at regional implementation of some key components of the Digital Health Framework (DHF) displayed in **Figure 3** of the document. The ultimate aim is to make the scheme depicted in **Figure 4** operational for different integrated care services, including CONNECARE case studies 1, 2 and 3.

Because of the existence of a well-defined regional risk assessment strategy, Lleida and Barcelona are jointly reported as one site regarding Task 2.3. A core strategic trait in the Catalan scenario is to foster convergence among: i) Evolution of existing regional assets; ii) Deployment initiatives at regional level (i.e. RIS3CAT initiatives like NEXTCARE - www.nextcarecat.cat); and, iii) Research and innovation projects. The three main regional assets considered for the purposes of the CONNECARE project are:

Adjusted Morbidity Groups (GMA) - The Catalan Health Surveillance System (CHSS) includes updated registries of the region of Catalonia (ES) (7.5M inhabitants) from Primary Care, Hospital-related events (hospitalizations, emergency room consultations and specialized outpatient visits), Pharmacy, Mental Health, Socio-sanitary services and other items (home-based respiratory therapies, dialysis, outpatient rehabilitation and non-urgent healthcare transportation) since 2011 [21,22]. It allows analyses of the use of healthcare resources, pharmacy consumption, prevalence of key disorders and population-based health risk assessment [1,20]. It is of note that although integration of CHSS registry data with electronic medical records is not yet in place, it constitutes the main goal of the PADRIS program [55], officially launched on January 2017.

The regional population-based health risk assessment tool, named GMA (Adjusted Morbidity Groups), is used to elaborate the health risk strata pyramid of the general population of Catalonia [1,20]. The GMA tool predicts individual patient risk, periodically updated on a six-month basis, based on multi-morbidity information gathered from CHSS registry data. The rationale behind the use of GMA, against alternative health risk assessment tools, is that it complies with four main recommended criteria described in **Table 1** [1], that is: (i) a population health approach (uses the entire population of 7.5M inhabitants of the region); (ii) without licensing constraints; (iii) open source computational algorithms; and, iv) the adjusted

morbidity grouper relies mostly on statistical criteria, as opposed to other tools that include expert-based coefficients, thus facilitating quick transferability to other territories. Detailed descriptions of the GMA, as well as its evaluation, have been reported elsewhere [20]. Unpublished results strongly support the core hypothesis of CONNECARE suggesting that GMA: i) Is a better index of multimorbidity-associated use of resources in primary care, as compared with other indicators (CRG, ACG, Charlson, etc…); and, ii) it has high potential to predict clinical events [20].

Overall, the GMA shows flexibility and transferability, as demonstrated by its recent adoption by thirteen out of the seventeen regional healthcare systems in Spain, covering 92% of the overall Spanish population, approximately 38 million citizens. Consequently, GMA can be considered a health risk assessment model that overcomes the main limitations identified in [1,20]. Accordingly, it seems suited for further assessment of the potential of population-health risk assessment and to enhance clinical risk modelling, as depicted in **Figure 1**.

Cat@Salut La Meva Salut (LMS) – The Personal Health Folder of Catalonia is linked to the Catalan interoperability system (shared electronic medical record at regional level, HC^3), and provides citizens with an access point to information about their health insurance. Cat@Salut LMS can also act as the citizen entry point for some of the supported processes (e.g. Medical appointments) and potentially for informal health data sources (e.g. mobile health applications, community medical devices, etc.). One of the aims of the project is to foster transformation of Cat@Salut LMS into a patient-self management tool supporting the three use cases of CONNECARE.

The regional interoperability framework - The Catalan health information exchange system (WiFiS) integrates basic highly standardized processes, namely: medical appointments, clinical data exchange, medical referral, etc., among healthcare providers with heterogeneous proprietary systems. It could also perform sectorial message routing and message delivery control. The shared electronic health record (HC^3) of Catalonia is a single system of medical records shared between different actors. The HC^3 enables the: i) Display of information that collects socio-demographic data of the citizen, documents or reports, prescriptions and immunization plus a summary screen with the most recent and relevant references; ii) provision of direct messaging between professionals to facilitate their cooperation; and iii) adds, at a later date, with ad hoc rules, clinical data provided by the private health sector or the proper citizen. The CONNECARE project is fully aligned with current technological developments on top of the current interoperability system aiming at supporting collaborative

adaptive case management, as described in deliverable *D2.2. Adaptive Case Management Design*, and depicted in **Figure 6**.

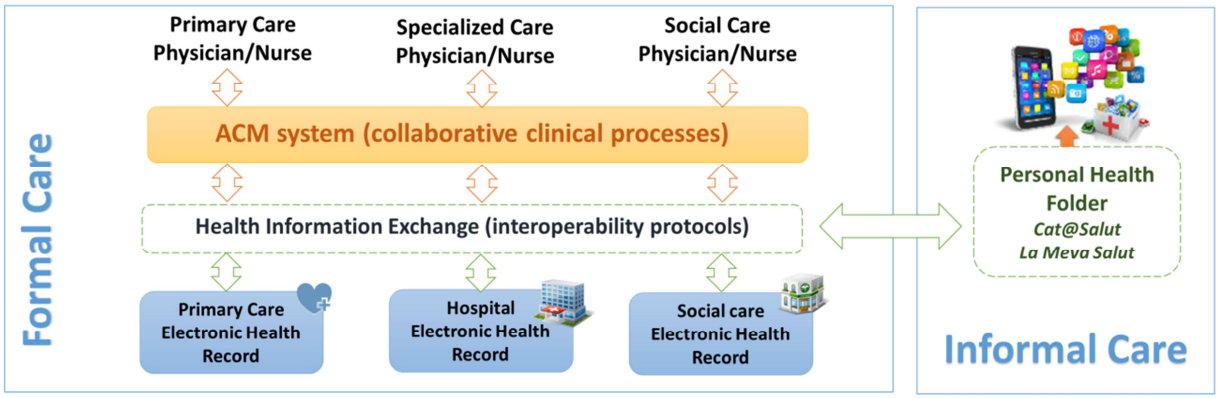


Figure 6 – Regional (Catalan) interoperability framework to support Adaptive Case Management (ACM) at a glance.

The general aims regarding health risk assessment are twofold:

1. To assess the potential of the Catalan Health Surveillance System (CHSS) and its population-based health risk assessment tool based on adjusted morbidity groups (GMA) to ***enhance clinical risk prediction***. The three use cases of CONNECARE will be the basis for elaboration and evaluation of two families of enhanced clinical predictive modelling, as briefly described below.

2. To elaborate a roadmap for use of multiple sources of information to refine **clinical risk prediction** for the individual patient. The ultimate aim is to support early diagnosis and preventive interventions, using: (1) Electronic Medical Records, (2) Registry data (CHSS & GMA), (3) Biomedical research info ("omics"), and, (4) Informal care data (Personal Health Folder) (**Figures 1-3**)

During the project life span, Lleida and Barcelona are planning to generate enhanced clinical health risk predictive modelling for the three use cases in order to fulfil a twofold aim: i) To assess the added value, in terms of both robustness and potential for model generalization, of incorporating refined GMA versions into patient-based risk assessment; and, ii) To feed clinical decision support systems (CDSS) that should facilitate optimal allocation of patients into the most appropriate integrated care services.

Use Case 1 – Management of Complex Chronic Patients (CCP) – CONNECARE will elaborate and evaluate predictive modelling to assess: i) risk of failure within the period of home hospitalization; ii) risk of early hospital re-admission after home hospitalization discharge (at 30 and 90 days after discharge); and, iii) clusters of patients for transitional care service selection.

Initial predictive modelling will be carried out using historical data from 2006 to 2015 described in [56]. This information will be enriched with data mining in electronic medical records and further evaluated incorporating the GMA grading of the patients as a covariate into the predictive modelling, as explained in detail in Annex I. We estimate that predictive modelling generated using enriched historical data (2006-2015) will be evaluated with prospective information from 2016. Moreover, the strategies developed in Task 2.3 are closely aligned with the activity performed in Task 3.4 in order to generate CDSS to be assessed within the CONNECARE life span.

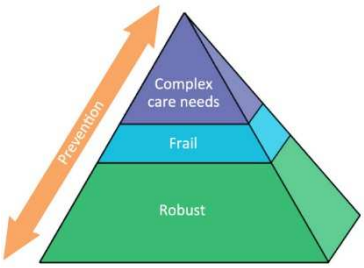
Use Cases 2 and 3 – Peri-surgical care – During the first year of the CONNECARE project, we have reported efficacy [57] and cost-effectiveness [58] of pre-habilitation in high risk candidates for major abdominal surgery. Prehabilitation (Case 3) has been adopted as a main stream service at Hospital Clinic. It is currently being extended to high risk candidates to other types of major surgery (cardiovascular, gynaecological, etc.). Moreover, we are in the process of designing a complete peri-operative program encompassing pre- and post-surgical care. De facto, we will be merging use case 2 and use case 3. Such a broad focus is generating requirements in terms of health risk prediction and allocation of candidates into appropriate peri-surgical care services with a community-based approach. We believe that the strategy for clinical risk prediction developed for case 1 will be generalized for cases 2 and 3 as soon as a proper evaluation of case 1 predictive modelling is completed. In summary, our estimation is that during the first quarter of 2018, we should be able to design the studies aiming at patient-based risk stratification of peri-surgical care.

2.2 Groningen (NL)

Case study 1: Embrace - Embrace is a person-centred and integrated primary care service that includes the population of all community living older adults aged 75-years or older. The development of Embrace is based on two evidence-based and wold-wide accepted models for organizing care and support, the Chronic Care Model and a population health management

model, the Kaiser Permanente Triangle. Both models were translated to the Dutch situation and specified for older adults. All structural aspects of both models were implemented at the same time and mutual connected.

Based on the Chronic Care Model (CCM), the health system is connected with community services, and reflects the four key elements of the CCM: self-management support, delivery system design, decision support, and clinical information systems. These are combined with a population health management model in terms of the Kaiser Permanente triangle to classify community-living older adults into risk profiles.



The formal risk stratification approach used in Embrace is based on the response of community living older adults on the annual (postal) questionnaire that participating older adult receive from their GP. This questionnaire contains, among other questions, the Intermed-SA-E and the Groningen Frailty Indicator for classification of the individual older adult into a risk profile.

Both on the level of the individual and the population risk is being stratified into three risk profiles according to levels of complexity of care needs and frailty. The risk profile “Robust” comprises participants without complex care needs and relatively low levels of frailty. These older adults hardly experience any consequences of aging and have an active social life, but are at risk of health-related problems due to the ageing process. The risk profile “Frail” comprises participants at risk of developing complex care needs and higher levels of frailty. These older adults increasingly suffer from the consequences of aging and experience growing dependency on others, while simultaneously their social network is shrinking. The risk profile “Complex care needs” comprises participants with complex care needs. These older adults are dependent on professional support for several of the underlying aspects due to the consequences of aging, and are at risk for referral to a hospital or nursing home.

Elements of case identification, selection and evaluation are incorporated. The degree of deployment of the stratification approach is done on a yearly basis. The approach is used based on clinical criteria and deployed as a predictive tool for future health related problems.

The regional health plans for citizens in our region involves on the one hand taking care of people in their own home environment for as long as possible, outside the walls of the hospital. On the other hand, for patients being discharged out of the hospital monitoring health status and preventing early readmission is pivotal. The development and implementation of the CONNECARE system could support these ambitions in the following ways. First, it will empower citizens by providing a platform which can be accessed and gives an up to date overview of current medical status and treatment plans. Also, it will improve communication not only with friends and family but ultimately with care professionals. By sharing health and/or treatment plans also over time early detection of declining health can be signalled and appropriate steps taken before citizens require complex care in the hospital. This will support integration of care over the entire care continuum and improve quality of life.

Second, for patients being discharged from the hospital the CONNECARE system will provide real-time information on current health status including vital signs. The system will allow more direct contact with care professionals in case of potential emergencies thereby filling up current gaps in communication following hospital discharge making it difficult for clinicians to monitor the well-being of patients. This will help streamlining conflicting advice given by different healthcare professionals involved in the recovery process. Being in direct contact will the care professional will also assist in the early detection of postoperative complications, an important reason for hospital readmission of surgical oncological patients. By applying early detection through the use of mobile devices complications and deviations from in postoperative recovery can be timely recognised and avoided.

Case study 1: The asthma and COPD Telehealth service - The asthma and COPD telehealth service is a cooperation between general practitioners, lung physicians and the department of general practice of the UMCG. The aim is to accomplish early diagnoses and treatment thereby enhancing asthma control, improving COPD health status, reducing morbidity, and improving the quality of life. This telehealth management support service assists GPs by examining patients and providing detailed diagnosis and treatment advice from pulmonologists through the internet. The aim is to provide an accurate and easily accessible service for GPs and patients, including also rural areas. Every patient suspected of having asthma, COPD, ACOS or who presents with pulmonary symptoms of unknown origin is eligible for inclusion.

Diagnosis and treatment advice based on a history questionnaire, the clinical COP Questionnaire, the Asthma Control Questionnaire and physical assessment in a laboratory

nearby. The data are filled in on a computer system by the laboratory assistant and are sent to the local pulmonologist by using an internet connection. The pulmonologist gives a working diagnosis and treatment advice through the computer and submits this to an online system. The GP receives this immediately through internet on his computer and can discuss the outcome with the patient.

The introduction of the CONNECARE system will expand the AC telehealth service in a couple of ways. First, the aim is to include the patient in the healthcare system thereby supporting disease management by accessing current information on health status. Also it will allow the patient to review the results of diagnostic tests and access to reliable websites on lung diseases, treatment and medication based on the advices of the pulmonologist. As such, a self-managed e-Health solution is offered in which the patient is in charge of his or own health plans. Self-management of patients is stimulated giving the patient a proactive role and central position in the decision making process. Tools will be provided within the CONNECARE system to empower patients, for instance by lifestyle monitoring through various mobile applications and devices. Changes in health status can be detected more easily and direct communication via chat with the professional will allow for prompt communication and adaptation of management plans when and if necessary. Patients can discuss the progress with their health care provider.

Case study 2: Major elective surgery - Within the patient cohort under investigation, patients are selected and stratified based on age (65 years and older), and classified as undergoing a high-risk surgical procedure for a solid malignant tumour in the operative centre of the University Medical Centre Groningen. High risk is defined as intracavitary surgery lasting more than 180 minutes. Care for these patients typically consists of a pre, peri, and post-operative phase. Currently, following discharge, the patient is scheduled for a two-week follow up at the outpatient clinic in the hospital. In between two or three phone consultations are planned. However, the degree of guidance and monitoring of recovery is limited once patients have been discharged from the hospital. The late diagnosis of post discharge complications has a negative impact on quality of life, clinical outcomes, medical consumption and increased healthcare costs. Postoperative care is a critical component of recovery following surgical procedures. More intensive monitoring of patients can increase speed of recovery, improve clinical outcomes, enhance patient satisfaction and decrease the rate of hospital readmissions. While clinic visits are the standard method of follow-up care, they can be non-practical at times due to patient immobility as a result of the nature of their surgery.

The CONNECARE system aims to further improve current disease management by using electronic data to enable patient monitoring, particularly in the postoperative setting. This approach could increase patient access to convenient postoperative care, expedite decisions about management, lead to earlier detection of postoperative complications, avoid unnecessary medical consumption and increase surgeon efficiency, making it a preferred method of follow-up for appropriate patients following surgery and discharge.

Therefore, improving the quality of postoperative care with e-health, with special focus on post-discharge care, may contribute to accelerated recovery, health care efficiency and better complication monitoring, which in turn may reduce readmission rates and health care costs. We hypothesize that with a novel smart, adaptive integrated care system to monitor elderly oncological patients after discharge following surgery, complications and deviations in postoperative recovery can be timely recognised and avoided.

2.3 Assuta (IL)

The Israeli healthcare System was one of the first internationally to begin computerizing its healthcare system. Maccabi Healthcare Services, the second largest Healthcare insurer/provider, was the pioneer in this area and began to computerize its physicians in 1989 – with all of their doctors working on electronic medical records by 1993. This was expanded to include nurses and all of the other health professionals within the following few years. In 1989, all Maccabi members received a magnetic membership card which is presented at each point of service such that ALL transactions with the patient (both clinical and administrative) are computerized and all of the data is stored in the Central data base. In 2001, Maccabi members were given access to their own medical information through a patient web portal and this can now be accessed by smartphone.

In addition to accessing his EMR data, the patient can access a full service directory and make appointments for all clinical services. The system has a GPS function that will show the patient services close to his current location. In addition, the web portal is now interactive and members can request prescription renewals and receive electronic prescriptions, likewise with referrals and other medical documents. All of this is captured in the Central Electronic Medical Record (EMR) and data base. As each member who uses the portal is registered in the system, tracking use of the portal also provides information on the technological level of the user as well as things such as whether he has WIFI at home, etc.

The above situation is similar in all 4 Israeli health plans. In addition, all hospitals in Israel are using electronic medical records and a system has been implemented enabling doctors in the community to see their patient's record while in hospital and vice-versa – the hospital doctor can see a patient's community EMR.

Maccabi and Assuta have taken this further as they are affiliated organizations. Maccabi has a doctor's web portal and Assuta physicians can access full information on their Maccabi patients via the portal. Maccabi and Assuta are taking the next step toward a fully integrated system with the opening of Assuta's new public hospital in the city of Ashdod, and this integrated system will be tested through the CONNECARE project. This will expand the clinical information in the data base with more specific hospital based data. In addition, it will expand the social services data as Assuta has reached an agreement with the Ashdod Social Services department on sharing data regarding hospitalized patients (with patient consent, of course).

Maccabi now has a longitudinal base containing over 20 years of patient data from birth to death on a stable population (< 2% annual turnover) of more than 2 million people. The data base is comprehensive, with extensive demographic data on each patient along with complete data on each patient from all clinical interactions and activities. The data base includes diagnoses and problems, medications prescribed and purchased, results of all diagnostic tests (laboratory, imaging, ECG and etc.), all visits to doctors and other health professionals, ER visits, hospitalizations. The data from the EMR has been enriched by administrative data and data from outside of the Maccabi system such as patients receiving personal caregiver services from the Social Security Institute, as well as patients who have applied for the service and not been accepted. In addition, as the full address of every member is in the database, we are able to correlate it with a Socio-Economic Status Analysis Using a GIS system. There is also geographically based socioeconomic data from the Central Bureau of Statistics and input into the Maccabi data base from location intelligence applications that analyse demographic characteristics in defined statistical geographic areas. Maccabi also has data on patients receiving welfare subsidies.

For purposes of patient stratification, Maccabi has organized its vast data base into patient registries including the medical problems depicted in **Table 4.** These patient registries are populated and updated daily using algorithms that run on the entire database. The registries form the basis for the Clinical Decision Support System (CDSS) which provides alerts and

recommendations to clinicians as well as to the patients themselves. Using advanced data mining techniques, it also supports predictive risk stratification.

Table 4. List of medical problems included in the patient registries.

Medical problems
Osteoporosis
Diabetes
Cardiovascular disease,
Infertility
Weight disorders
Hypertension
Schizophrenia/bipolar
Cancer
Warfarin Treatment
Chronic Obstructive Lung disease
Home Care
Complex patients

Examples of this are the identification of members that need to do prevention such as women who need to have a breast check and do mammography, members who need to get a flu shot, members who need to get a pneumovax vaccination and etc. The result of this appears as personal recommendations to the member – both on the member portal as well as proactively as a text message on their mobile phones, as well as pop us recommendations to the member's family doctor.

As an example of a more sophisticated risk assessment model, Maccabi together with Medial Research (a pioneer in the field of algorithmic analysis of medical data) developed and validated a model to identify individuals at increased risk for colorectal cancer (CRC) by analysing blood counts, age, and sex, and then determined the model's value when used to supplement conventional screening in a bi-national Israel-UK study.

Using blood counts obtained 3-6 months before diagnosis, the area under the curve for detecting CRC was 0.82 ± 0.01 for the Israeli validation set. The specificity was 88 ± 2% in the Israeli validation set and 94 ± 1% in the UK dataset. Detecting 50% of CRC cases, the odds ratio was 26 ± 5 and 40 ± 6, respectively, for a false-positive rate of 0.5%. Specificity for 50%

detection was 87 ± 2% a year before diagnosis and 85 ± 2% for localized cancers. When used in addition to the faecal occult blood test, our model enabled more than a 2-fold increase in CRC detection. This has now been incorporated into the Maccabi CDSS and clinicians are receiving alerts identifying individuals requiring additional clinical evaluation in order to detect CRC earlier in clinical practice [59].

Maccabi's complex patient registry is an example of an amalgam of population based risk stratification with patient based stratification given that the data based is comprised of all Maccabi members, the majority of whom are not necessarily patients. As noted above, the data base contains data from population based sources outside of the electronic medical record data. The purpose of the Registry is to identify Maccabi members at risk for potential exacerbation of their condition, higher use of healthcare resources, including hospitalization.

The approach pioneered in the development of the complex patient registry is guiding the integrated care system being jointly developed for complex chronic patients (CCP) by Assuta and Maccabi in Ashdod that includes a risk stratification model and algorithm used to identify patients for inclusion in the CCP registry. It is being implemented in the screen of the professional work station to identify patients at the time of hospitalization who require inpatient case management and follow-up integrated care in the community. The aim is to assure more rapid recovery and recuperation and to prevent excessive use of healthcare resources and preventable readmissions. This risk stratification model will be implemented in the CONNECARE project to identify patients for inclusion in both Case 1 (patients with an unplanned hospital admission discharged back to the community) and Case 2. (Chronically ill patients scheduled for major elective surgery who are expected to be discharged back to the community)

Case Study 1 – Management of Complex Patients in the Community

The CONNECARE project will assess the effectiveness of case management in the hospital, integration with the community and follow up integrated care in the community post discharge. Patients will be assessed and stratified for risk at 2 levels: Level 1 will take place In the Emergency department or immediately upon admission, where patients will be stratified using

risk stratification model and algorithm used to identify patients for inclusion in the CCP registry as described above, to identify patients in need of inhospital case management. Level 2 will take place during the hospitalization prior to discharge to identify those patients in need of follow up integrated care post discharge. These will be the patients recruited for

CONNECARE Case 1. A further level of stratification will take place to determine which patients are discharged to home hospitalization, to home rehabilitation, to home care supervised by the home care unit, or to home with his regular support system.

Case Study 2 – Management of Complex Patients Scheduled for Major Elective Surgery

Patients will be assessed and risk-stratified by the Surgical Department (General Surgery, Orthopedic surgery, Gynecological surgery) that will be performing the surgery at the point where the decision is made to perform the surgery. The referral to the surgical department will already identify the patient as having been stratified by Maccabi as a patient in the Complex patient registry. The surgeon will then assess the patient in accordance with the CONNECARE inclusion criteria – specifially classifying the patient according to ASA level. These patients will be recruited to CONNECARE Case Study 2 and will be given a prehabilitation program prior to surgery. Patients' compliance with the prehabilitation program will be monitored by the CONNECARE Self management app. Prior to surgery patients will be further stratified using the INTERRAI geriatric evaluation tool to determine which patients are at higher risk for post-surgical complications and these will be assigned an in-hospital case manager. Stratification to the various post discharge options will take place similarly to Case 1

3. CONNECARE proposals

3.1 Expected scenario for health risk assessment at the project end

Enhanced clinical risk assessment (Figure 7 # 3) is one out of the five intertwined dimensions targeted in CONNECARE to generate relevant research & innovation outcomes throughout the life span of the project, as indicated in **Figure 7**. The other four dimensions are: i) Clinical studies (# 1) carried out on top of large scale implementation of integrated care services for each of the uses cases in the four sites (WP6); ii) ICT developments (# 2) supporting the Chronic Care Model (CCM) (WPs 3-5); iii) Innovative evaluation strategies using a Triple Aim approach [60,61] (# 4) to be deployed during and beyond the project lifetime (WP7); and, iv) Final recommendations for regional adoption of the integrated care services (# 5), as well as transferability to other regions (WP7).

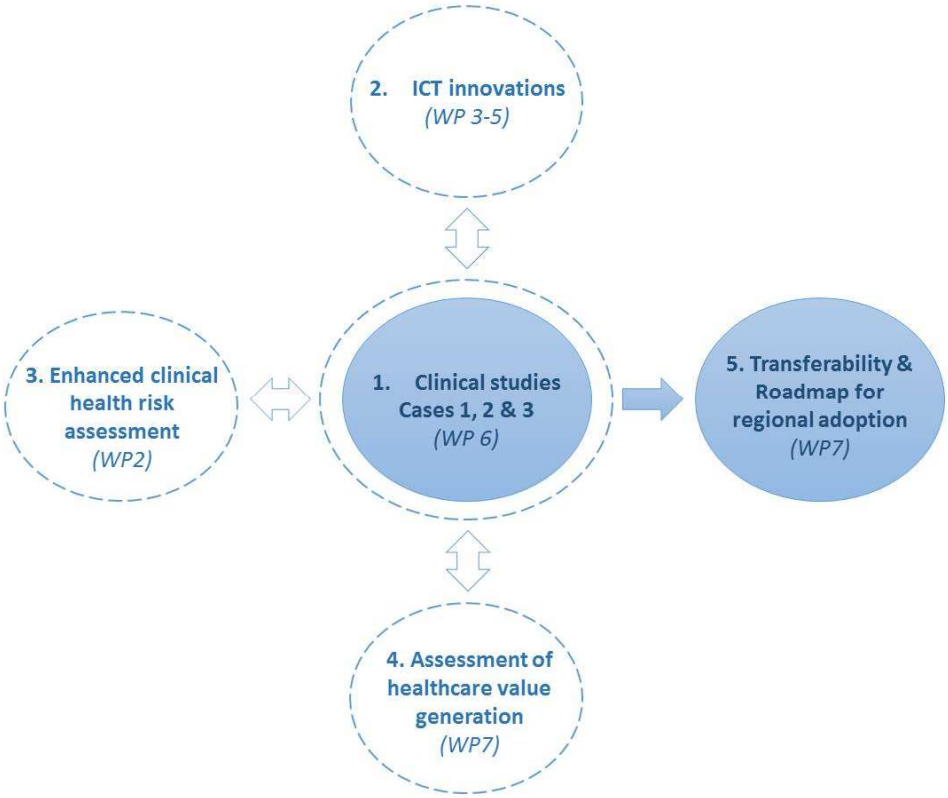


Figure 7 – Five pivotal aims of the CONNECARE project aiming at achieving successful regional adoption of the community-based protocol for collaborative management of complex chronic patients across health-care tiers [62]. The figure indicates allocation of the different work packages (WP) into each of the five aims of the project. The current document provides the frame for the protocol of the clinical studies associated to Case 1 (Aim 1). The current

working document will be enriched with assessment protocols for Aims 2 (ICT products); 3 (Risk Assessment tools); and, 4 (Overall assessment strategy following an Implementation Research approach). Moreover, the final outcome of the project will be generation of recommendations for Aim 5 (Transferability & Roadmap for regional adoption).

During the initial sixteen month-period of the project, the consortium has achieved a high degree of consensus regarding the working plan produced in Task 2.3. Briefly, we all share a common goal regarding both modalities and role of patient-based risk assessment in the clinical scenario at the project end, as described below. We acknowledge, however, that the degree of maturity of its implementation is different among the three areas: Catalonia (ES) - Lleida and Barcelona -; Northern Netherlands - Groningen (NL); and, Assuta (IL). Consequently, the PDSA strategies, see *D2.4. Case studies description and the associated co-design process* and the current document, will be highly site dependent.

Enhanced health risk assessment in the clinical setting

CONNECARE will set-up the new clinical predictive modelling approach based on the principles reported in **Sections 1.1** and **1.2** of the current document and it will be assessed in the case studies deployed in WP6.

The four main novelties of the enhanced clinical risk assessment strategy envisaged in CONNECARE are: i) Enrichment of clinical predictive modelling with outcomes from population-health predictive modelling; that is, inclusion of the patient's GMA grading as a covariate; ii) Inclusion of multilevel data sources in the predictive modelling; specifically, information from informal care sources; iii) Implementing ICT support for dynamic predictive modelling which should allow adaptive changes of the patient action plan over time based on updated risk assessment; and, iv) Feeding clinical decision support systems (CDSS) with outcomes of predictive modelling such that decisions of health professionals across the structured service workflow, or patient work plan, could be facilitated using information generated by clinical risk predictive modelling.

The scheme depicted in **Figure 4** illustrates how the CONNECARE personalized self-management system (WP4), implemented into the collaborative adaptive care management plan, will enrich clinical information with informal care data, the latter channelled by the personal health folder, that will be contributing to dynamically feed, through predictive

modelling tools, the different CDSSs *(WP3 – D 3.2: First screening and risk stratification DSS)* supporting health professional in the decision making processes.

Expected outcomes at the project end

Specific outcomes at the end of the project for each of the four novelty items generated in Task 2.3 are reported below:

1. **Enrichment of clinical predictive modelling with inclusion of patient's GMA grading as a covariate:**
   - ✓ Generation of enriched predictive modelling and evaluation for complex chronic patients in Catalonia will be completed within 2018.
   - ✓ Analysis of transferability of the GMA tool to other sites (IL and likely NL) will be completed throughout the project life time.
2. **Inclusion of multilevel data sources in the predictive modelling; specifically, information from informal care sources:**
   - ✓ It will be addressed in Catalonia through current transformation of the personal health folder (La Meva Salut) during 2018 and completed during the project lifespan.
   - ✓ Specific modalities for multilevel data inclusion will be assessed in IL (InterRAI) and NL.
3. **Implementing ICT support for dynamic predictive modelling:**
   - ✓ Testing of the CONNECARE personalized self-management system (WP4) for this purpose will begin at the end of 2017 and it will be completed at M36 having a prototype with an expected TRL (Technology Readiness Level) = 6 regarding this specific Task 2.3 item.
4. **Feeding CDSSs with outcomes of predictive modelling:**
   - ✓ Testing of the CDSSs (Task 3.4) included into the CONNECARE Smart Adaptive Case Management platform (WP3) will be completed at M36 having a prototype with an expected TRL= 6 regarding this specific Task 2.3 item.
   - ✓ Groningen envisages the development of a roadmap for coupling the CONNECARE system to the electronic patient dossier described below.
5. **Site specific roadmap toward a full implementation of the CONNECARE enhanced clinical risk assessment.** It is assumed that the three areas (ES, NL and IL) will still show heterogeneities regarding Task 2.3 at the end of the project. Consequently, each of the areas will produce a roadmap describing progression beyond the project in order to bring items 2 to 4 to a mature TRL equal to 9.

The CONNECARE work plan described in Task 2.3 will address the patient-based five-dimension scoring strategy generated within the project, consisting in (i) Screening for ease

CCP identification; (ii) Risk stratification for hospital admission and emergency room frequentation; (iii) Mapping of complex chronic patients (CCP) and clinical and social issues; (iv) Interventions structured in a proactive and planned work plan based on risk stratification and mapping; and (v) Continuous surveillance of changes in clinical and social status of CCP involving changes at risk stratification.

3.2 The need for specific PDSA approaches to health risk assessment in each site

Specific PDSA strategies are currently being elaborated in each site in order to achieve a smooth transition from the current initial status to the commitments defined for the end of the project and beyond, as described in the previous section. The methodology of the PDSA cycles is the one already described in *D2.4. Case studies description and the associated co-design process*. Its site specificities, as well as the PDSA outcomes of Task 2.3 achieved during the initial twenty-month period of CONNECARE, will be reported in *D7.2. Evaluation results of the initial co-design phase until Study Release*.

Within this subheading, we are briefly describing the steps undertaken in each of the three areas (ES, IL and NL) regarding Task 2.3.

Catalonia (ES): Barcelona and Lleida

The elaboration of predictive modelling for CCP is being carried out using historical data, from 2006 to 2015, of the Home Hospitalization and Early Discharge (HH/ED) service at Hospital Clinic de Barcelona (HCB-IDIBAPS) [56]. The original dataset will be enriched with additional information obtained mining electronica health records (EHR) and also testing the added value of including the GMA grading as a covariate. The predictive modelling resulting from these analyses will be evaluated using 2016 data. The developed predictive models will be integrated in the CONNECARE CDSS system as a service for assisting HH/ED decision making. A summary of the protocol is described in **Annex I**.

Groningen (NL)

The University Medical Centre Groningen is currently implementing an electronic patient dossier (EPD) making it possible to further connect information systems for patients. An important goal is to place the patient at the centre of information and communication through an easy to use platform. As current EPD hospitals systems are primarily developed for the professional and not the patient, the CONNECARE system will be ideally placed to empower patients and put them in the centre of the health care system. However, change is not expected to occur easily and has to be part of short improvement cycles in order to involve all relevant stakeholders and to create ownership among clinicians, policy makers and IT personnel. Ultimately connecting the CONNECARE IT infrastructure with the EPD will allow for developing personalized risk assessment models which can be fed back into the CONNECARE system of the patient.

Assuta (IL)

The CONNECARE project provides Assuta and Maccabi a unique opportunity to enrich and expand their capacity for predictive health risk assessment. Given that Maccabi and Assuta have expanded their data bases to include data beyond EHR data, as described in Section 2, as part of the CONNECARE project, we plan to explore the possibility of applying the GMA Population health algorithm developed in Spain to our data base in order to determine whether the combination of the current risk assessment model will be enriched by the GMA population based risk assessment and achieve a more precise predictive model, and to what extent it may lead to more focused, personalized risk assessment and prevention (see preliminary protocol described in **Annex II**).

4. Conclusions

The current document reflects the consensus of the consortium regarding the principles governing the enhanced clinical risk assessment strategies to be implemented into CONNECARE (Sections 1.1 and 1.2). Section 2 reports on the status of the three areas (ES, NL and IL) regarding maturity of risk assessment, as well as information on the clinical studies. Finally, Finally, It is of note, that Lleida and Barcelona have been merged as one area regarding Task 2.3 because the risk assessment strategy is highly governed at regional level through specific policies promoting interaction between the Catalan population-health predictive modelling tool (GMA, Adjusted Morbidity Groups) and clinical risk prediction. Section 3 indicates the common goals to be achieved at the end of the project and provides indications on the PDSA strategies while indicating the initial steps already activated in each site.

5. References

1 Dueñas-Espín I, Vela E, Pauws S, *et al.* Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* 2016;**6**:e010301. doi:10.1136/bmjopen-2015-010301

2 Maddox TM, Albert NM, Borden WB, *et al.* The Learning Healthcare System and Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circulation* 2017.

3 Bousquet J, Anto JM, Sterk PJ, *et al.* Systems medicine and integrated care to combat chronic noncommunicable diseases. *Genome Med* 2011;**3**:43. doi:10.1186/gm259

4 Sansone SA, Rocca-Serra P, Field D, *et al.* Toward interoperable bioscience data. *Nat Genet* 2012;**44**:121–6. doi:10.1038/ng.1054

5 Barabási A-L. Network Medicine — From Obesity to the 'Diseasome'. N. Engl. J. Med. 2007;**357**:404–7. doi:10.1056/NEJMe078114

6 Fujita KA, Ostaszewski M, Matsuoka Y, *et al.* Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. *Mol Neurobiol* 2014;**49**:88–102. doi:10.1007/s12035-013-8489-4

7 Bhinder B, Elemento O. Towards a better cancer precision medicine: Systems biology meets immunotherapy. *Curr Opin Syst Biol* 2017;**2**:67–73. doi:10.1016/j.coisb.2017.01.006

8 Cano I, Lluch-Ariet M, Gomez-Cabrero D, *et al.* Biomedical research in a Digital Health Framework. *J Transl Med* 2014;**12**:S10. doi:10.1186/1479-5876-12-S2-S10

9 Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. *IEEE J Biomed Heal informatics* 2015;**19**:1209–15. doi:10.1109/JBHI.2015.2406883

10 Andreu-Perez J, Poon CCY, Merrifield RD, *et al.* Big Data for Health. IEEE J. Biomed. Heal. Informatics. 2015;**19**:1193–208. doi:10.1109/JBHI.2015.2450362

11 Hamad R, Modrek S, Kubo J, *et al.* Using 'big data' to capture overall health status: properties and predictive value of a claims-based health risk score. *PLoS One* 2015;**10**:e0126054. doi:10.1371/journal.pone.0126054

12 Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;**309**:1351–2. doi:10.1001/jama.2013.393

13 Belle A, Thiagarajan R, Soroushmehr SMR, *et al.* Big Data Analytics in Healthcare. *Biomed Res Int* 2015;**2015**:370194. doi:10.1155/2015/370194

14 Luo J, Wu M, Gopukumar D, *et al.* Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights* 2016;**8**:1–10. doi:10.4137/BII.S31559

15 Gerstein M. Genomics: ENCODE leads the way on big data. *Nature* 2012;**489**:208–208. doi:10.1038/489208b

16 Robertson ARR, Nurmatov U, Sood HS, *et al.* A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data. *J Innov Heal informatics* 2016;**23**:611–9.

17 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36. doi:10.1093/ije/dyv098

18 McDonald L, Lambrelli D, Wasiak R, *et al.* Real-world data in the United Kingdom: opportunities and challenges. *BMC Med* 2016;**14**:97. doi:10.1186/s12916-016-0647-x

19 WHO. Innovative Care for Chronic Conditions: Building Blocks for Action. Geneva: World Health Organization (WHO/MNC/CCH/02.01). http://www.who.int/chp/knowledge/publications/icccglobalreport.pdf. Date last accessed: March 9 2017. 2002.

20 Monterde D, Vela E, Clèries M. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Atención Primaria* 2016;**48**:674–82. doi:10.1016/j.aprim.2016.06.003

21 Farré N, Vela E, Clèries M, *et al.* Real world heart failure epidemiology and outcome: A population-based analysis of 88,195 patients. *PLoS One* 2017;**12**:e0172745. doi:10.1371/journal.pone.0172745

22 Farré N, Vela E, Clèries M, *et al.* Medical resource use and expenditure in patients with chronic heart failure: a population-based analysis of 88 195 patients. *Eur J Heart Fail* 2016;**18**:1132–40. doi:10.1002/ejhf.549

23 Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med* 2006;**355**:2615–7. doi:10.1056/NEJMp068249

24 Kindig D, Stoddart G. What is population health? *Am J Public Health* 2003;**93**:380–3. doi:10.2105/AJPH.93.3.380

25 Kindig D. What Are We Talking About When We Talk About Population Health? – Health Affairs Blog. 2015.

26 Kansagara D, Englander H, Salanitro A, *et al.* Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;**306**:1688–98. doi:10.1001/jama.2011.1515

27 Roland M, Abel G. Reducing emergency admissions: are we on the right track? *BMJ* 2012;**345**.

28 Johnson TL, Rinehart DJ, Durfee J, *et al.* For Many Patients Who Use Large Amounts Of Health Care Services, The Need Is Intense Yet Temporary. *Health Aff* 2015;**34**:1312–9. doi:10.1377/hlthaff.2014.1186

29 Lewis G, Curry N, Bardsley M. Choosing a predictive risk model : a guide for commissioners in England. 2011;:20.

30 NIGB National Information Governance Board. Advice on Risk Prediction and Stratification Activities. 2012.

31 Lewis GH. 'Impactibility models': Identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q* 2010;**88**:240–55. doi:10.1111/j.1468-0009.2010.00597.x

32 Orkin FK. Risk stratification, risk adjustment, and other risks. *Anesthesiology* 2010;**113**:1001–3. doi:10.1097/ALN.0b013e3181f7ab17

33 JW T. Risk Adjustment for Measuring Health Care Outcomes, 3rd edition. *Int J Qual Heal Care* 2004;**16**:181.

34 Mora J, Iturralde MD, Prieto L, *et al.* Key aspects related to implementation of risk stratification in health care systems-the ASSEHS study. *BMC Health Serv Res* 2017;**17**:331. doi:10.1186/s12913-017-2275-3

35 Esteban de Manuel Keenoy, Marco Nalin, Tamara Alhambra, Francesca Avolio IB, Anna Bedbrook, Barbara Branchini, Joan Carlos Contel Segura, Miren David Iturralde D, De Massari, Cristina Domingo Rico, Irati Erreguerena, Santiago Esnaola Sukia JG, *et al.* White paper on Deployment of Stratification Methods. The ASSEHS project.

36 Advancing Care Coordination and TeleHealth (ACT). https://www.act-at-scale.eu/

37 Health and Social Care Programme Scotland. Scottish Patients at Risk of Readmission and Admission (SPARRA) A Report on the Development of SPARRA Version 3 - Developing Risk Prediction to Support pReventive and Antcipatory Care in Scotland. vol. 3. 2011.

38 Wharam JF, Weiner JP. The promise and peril of healthcare forecasting. *Am J Manag Care* 2012;**18**:e82-5.

39 Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. World Health Organization - Europe, European Observatory on Health Systems and Policies, Cambridge; 2009. doi:10.1017/CBO9780511711800

40 Weiner JP, Trish E, Abrams C, *et al.* Adjusting for risk selection in state health insurance exchanges will be critically important and feasible, but not easy. *Health Aff* 2012;**31**:306–15. doi:10.1377/hlthaff.2011.0420

41 Vela E, Tényi Á, Cano I, *et al.* Population-based analysis of COPD patients in Catalonia: implications for case management. *BMJ Open* 2017;**2nd revisi**.

42 Coughlin SS. Toward a road map for global -omics: a primer on -omic technologies. *Am J Epidemiol* 2014;**180**:1188–95. doi:10.1093/aje/kwu262

43 Cano I, Tenyi A, Vela E, *et al.* Perspectives on Big Data applications of health information. *Curr Opin Syst Biol* 2017;**3**:36–42. doi:10.1016/j.coisb.2017.04.012

44 Big Data in Public Health, Telemedicine and Health. https://ec.europa.eu/digital-single-market/en/news/study-big-data-public-health-telemedicine-and-healthcare

45 Gomez-Cabrero D, Menche J, Vargas C, *et al.* From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformatics* 2016;**17**:23–35. doi:10.1186/s12859-016-1291-3

46 Gallego B, Walter SR, Day RO, *et al.* Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. *J Comp Eff Res* 2015;**4**:191–7. doi:10.2217/cer.15.12

47 Brown S-A. Patient Similarity: Emerging Concepts in Systems and Precision Medicine. *Front Physiol* 2016;**7**:561. doi:10.3389/fphys.2016.00561

48 Haas LR, Takahashi PY, Shah ND, *et al.* Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care* 2013;**19**:725–32.

49 Purdy S, Huntley A. Predicting and preventing avoidable hospital admissions: A review. *J R Coll Physicians Edinb* 2013;**43**:340–4. doi:10.4997/JRCPE.2013.415

50 Sharafoddini A, Dubin JA, Lee J. Patient Similarity in Prediction Models Based on Health Data: A Scoping Review. *JMIR Med informatics* 2017;**5**:e7. doi:10.2196/medinform.6730

51 Henriques J, Carvalho P, Paredes S, *et al.* Prediction of Heart Failure Decompensation Events by Trend Analysis of Telemonitoring Data. *IEEE J Biomed Heal Informatics* 2015;**19**:1757–69. doi:10.1109/JBHI.2014.2358715

52 Lee J, Maslove DM, Dubin JA, *et al.* Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric. *PLoS One* 2015;**10**:e0127428. doi:10.1371/journal.pone.0127428

53 Wang Y, Tian Y, Tian L-L, *et al.* An electronic medical record system with treatment recommendations based on patient similarity. *J Med Syst* 2015;**39**:55. doi:10.1007/s10916-015-0237-z

54 MD. S. Best Care at Lower Cost: The Path to Continously Learning Health Care in America.

55 Department of Health C. Programa públic d'analítica de dades per a la recerca i la innovació en salut (PADRIS).

http://salutweb.gencat.cat/web/.content/home/ambits_tematics/linies_dactuacio/recerca/enllacos/Programa_analitica_dades_PADRIS_aquas2017_publica.pdf. Date last accesse. 2017.

56 Hernández C, Alonso A, Garcia-Aymerich J, *et al.* Integrated care services: lessons learned from the deployment of the NEXES project. *Int J Integr Care* 2015;**15**:e006. doi:10.5334/ijic.2018

57 Barberan-Garcia A, Ubré M, Roca J, *et al.* Personalised Prehabilitation in High-risk Patients Undergoing Elective Major Abdominal Surgery. *Ann Surg* 2017;:1. doi:10.1097/SLA.0000000000002293

58 Barberan-Garcia A et al. Cost-efficacy analysis of Personalised Prehabilitation in High-risk Patients Undergoing Elective Major Abdominal Surgery.

59 Kinar Y, Kalkstein N, Akiva P, *et al.* Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *J Am Med Informatics Assoc* 2016;**23**:879–90. doi:10.1093/jamia/ocv195

60 Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. *Health Aff* 2008;**27**:759–69. doi:10.1377/hlthaff.27.3.759

61 WHITTINGTON JW, NOLAN K, LEWIS N, *et al.* Pursuing the Triple Aim: The First 7 Years. *Milbank Q* 2015;**93**:263–300. doi:10.1111/1468-0009.12122

62 Cano I, Dueñas-Espín I, Hernandez C, *et al.* Protocol for regional implementation of community-based collaborative management of complex chronic patients. *npj Prim Care Respir Med* 2017;**27**:44. doi:10.1038/s41533-017-0043-9

6. Annex I - Current developments for use case 1 in Barcelona

At the Barcelona site, the goal for use case 1 is to develop predictive risk models that can help in the home hospitalization service to aid the service selection of CCP patients. The hypothesis is that predictive modelling using clinical data could be significantly improved by enriching computational models with covariates reflecting outcomes from population-health risk prediction and information extracted from EMRs, as explained in detail in section 2.

Aims - To develop and validate enhanced clinical predictive modelling for HH/ED with a two-fold aim:

- ✓ During the HH/ED period (t0) - To identify risk of early readmission after hospital discharge and mortality to stratify patients in order to optimize care (RM1-3).
- ✓ After HH/ED discharge (t1) - To identify risk of early readmission after hospital discharge and mortality and stratify patients for transitional care purposes (RM4-6).

Figure 1 in the current Annex summarises the plans for the predictive modelling for the home hospitalization program together with the data that is available at each time point. In the study, early readmissions were divided into groups of emergency cases (Readmission ER) and simple early readmission cases (Hospitalization).

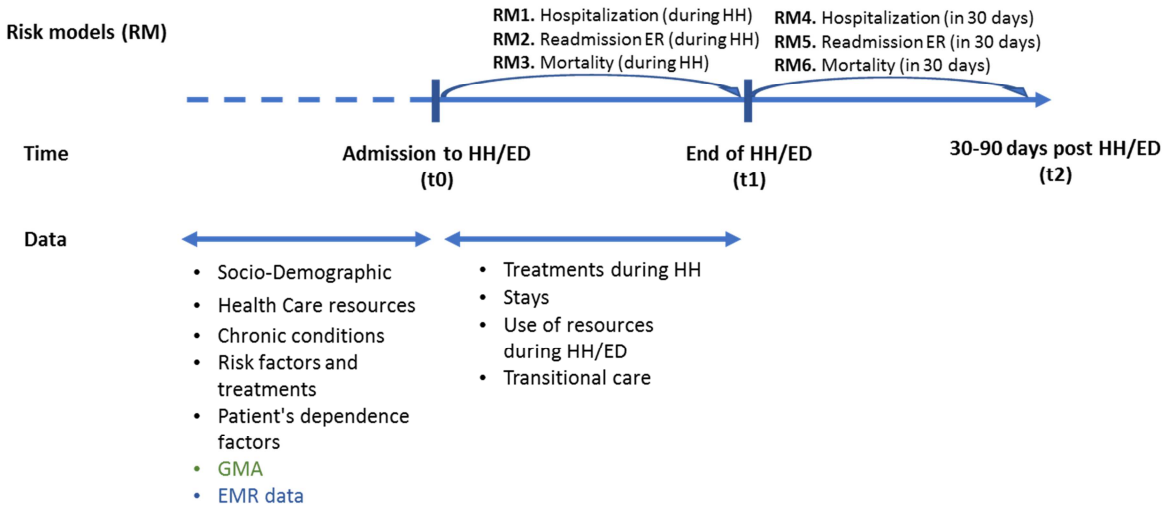


Figure 1 *- Schematics of the plans for the predictive modelling.*

Data extraction strategy

The ultimate aim of the data extraction strategy is to integrate several data sources containing health related patient information, namely:

1) Registry data (CHSS & GMA),
2) Electronic Medical Records (SAP EMR, eCAP EMR),
3) *Biomedical research data ("omics"),*
4) *Informal care data (self-management and lifestyle via personal health folders).*

Currently, only the collection of registry and EMR data is possible but in future the biomedical and informal care data should be also expected to contribute to risk modelling. Initially a specific database generated during the HH/ED program will be used for modelling purposes. This database contains patient information on socio-demographics, risk factors, main diagnosis, treatment-related data, etc. (see **Figure 1**). Secondly, GMA scores for population-based risk prediction will be integrated from static data source in the first round, however future plans include its direct extraction from the Catalan Health Surveillance System (CHSS). Further steps will also consider the integration of EMR data extracted from the hospital's SAP system and the eCAP primary care EMR system in the frame of the PADRIS program [55].

Modelling strategy

The aims of the modelling strategy are two-fold (**Figure 2**).

- ✓ Evaluate the additional predictive power introduced by GMA and EMR data (mostly the ones extracted from SAP) compared to existing population-based models (Hernandez C et al. 2017; Vela E et al 2017, both papers are submitted for publication),
- ✓ Increase predictive performance of the models by partitioning the population into smaller, biologically similar groups of patients.

The first aim of the modelling is to validate the impact of the increased information content gained through the data extraction as compared to previous studies. Logistic regression modelling is used to recreate the model (Hernandez C et al. 2017, submitted for publication) and the area under the ROC curve (AUC) is used to evaluate its performance. Then, GMA as

a covariate is added to the model to evaluate the difference in AUC. Same methodology is followed using EMR data.

The population included in the home hospitalization is known to be heterogeneous in terms of main diagnosis and clinical manifestation. Therefore, the second aim is to increase predictive performance of the models, by partitioning the population into smaller, similar groups of patients. Patient similarity-based clustering is used to generate virtual patient cohorts (clusters). Similarity is computed using Gower distance measure on covariates, which were selected using a prior model computed using a step-wise variable selection method on the entire study population. For each group, then models are generated with the same step-wise variable selection method, their predictive performance is evaluated and group specific risk factor profiles are retrieved.

Finally, this approach is compared to Fuzzy c-Regression Models (FCRM), which is an algorithm that over the steps mentioned above also performs a regression-based optimization of the final cluster generation and risk prediction. The latter algorithm should function as our recommended methodology for risk stratification.

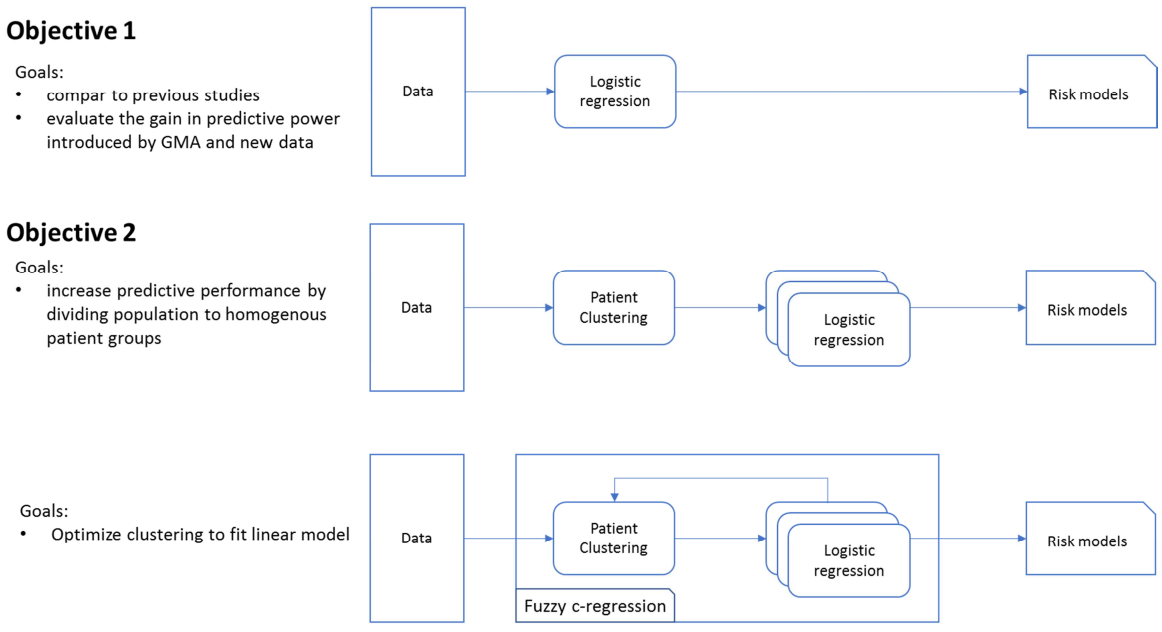


Figure 2 *- Modelling strategy.*

Deployment strategy

The final models shall inform on patients risk on readmission and mortality. In CONNECARE these models will be plugged into the CDSS system developed in Task 3.4. Following a collaborative approach with the team implementing the CDSS system, we put a great emphasis on making these two tasks interoperable.

Main barriers to integrate the developed models in the CDSS system are organisational as well as technical:

- On the one hand, predictive models are usually developed by data scientists or statisticians often using non-open data and closely collaborating with healthcare practitioners to ensure thorough validation and widespread agreement on their exploitation. Therefore, externally developed models might be needed to be integrated in the system.
- On the other hand, data scientists and statisticians usually rely on domain specific programming languages and frameworks such as the R language[1] or Pyhton's scikit-learn[2] to develop their predictive models. These languages may not be adequate for developing a general purpose, distributed, modular, and web-based system as CONNECARE is, besides not being the ones actually used in developing the CONNECARE system by the technical partners of the project – that is, mainly Java[3].

The solution to overcome these issues proposed by CONNECARE leverages the existing PMML/PFA standard formats for representing machine learning and statistical (already trained) models[4] – with the latter supporting also full machine learning pipelines and dataflows. The PMML/PFA standards are XML/Json -based representations of predictive models – such as regression models, classifiers, clustering methods, etc.–, which allow for seamless exchange of models ready to be exploited for making predictions across machine learning / statistical toolkits and languages. The main idea is to allow for separate development and deployment of predictive models, ensuring the flexibility of the production environment to easily integrate novel models nevertheless its source, i.e. a data science team collaborating with healthcare practitioners or the analytics team developing the system. Using

[1] https://www.r-project.org/about.html

[2] http://scikit-learn.org/stable/

[3] https://www.oracle.com/java/index.html

[4] http://dmg.org/

such flexible standards also allow for the development of models in different programming languages, e.g. R and Python, while they can be integrated into the production environment, such as a Java-based RESTful web service, once they have been translated into PMML.

In the specific context of CONNECARE, the CDSS will provide a RESTful endpoint exposing API to published/unpublished models, apply them to new data, query their parameters, further train them, and similar, seamlessly operating with R and Python models, and possibly any other model once a suitable translation module has been developed. This way, clinical partners in each site are free to develop their own models in their own preferred analytics environment and leveraging their own (possibly, private) data, while the CDSS takes care of transparently translating them into PMML/PFA format, ready to be applied on novel data, shared with other sites, and many other functionalities.

This solution will be thoroughly discussed in the upcoming deliverable on First Screening and Risk Stratification DSS (D3.4).

7. Annex II - Planned developments for use case 1 and 2 in Israel

The aim of the operation in the IL site twofold:

1. To develop and implement a risk assessment model based on the totality of data available in the Maccabi and Assuta data bases (the Maccabi data base contains over 20 years of comprehensive data on over 2 million Maccabi members)in order to identify patients who require special case management in the hospital and intensive follow-up case management post-discharge (both CONNECARE Case 1 and Case 2)
2. To apply the GMA population-health algorithm developed in Spain to our data base in order to determine whether the GMA population-based risk assessment model can achieve a more precise predictive model, and to what extent it may lead to more focused, personalized risk assessment and prevention.

In order to achieve the integration of the GMA method in the system of IL site the following steps will be taken:

1. To achieve the first usable version for exploratory purposes, the current system trained on the Catalan population will be implemented.
2. For final integration, the method parameters will be updated to reflect the specificities of the IL site population.

To achieve the first implementation, all the patient diagnoses (coded with ICD-9-CM) will be retrieved from the EMR system of IL. This allows for the stratification of the population according to the GMA groups and evaluation of personal GMA scores of the patients. This will be used to run first test, integrating it in clinical models and evaluate improvements.

In the second step, data on mortality and healthcare service utilization (primary care, hospitalization and pharmacy) will be also retrieved from the EMR system. This will be used to update the parameters of the GMA model with the specificities of the population of the site, for more personalized predictions.