AHRQ R21 - Individualized Precision Prevention - Report

Title Page

Title of Project.
Achieving Individualized Precision Prevention (IPP) through Scalable Infrastructure Employing the USPSTF Recommendations in Computable Form.

Principal Investigator and Team Members.
Dr. Charles P. Friedman, Principal Investigator
Dr. Tanner Caverly, Investigator
Dr. Glen Taksler, Investigator
Dr. Allen Flynn, Investigator

Organization.
University of Michigan

Inclusive Dates of Project
7/1/2018 - 6/30/2020

Federal Project Officer
Derrick Wyatt, MSN, RN-BC, CPHIMS

Acknowledgment of Agency Support.
Supported by grant number 1 R21 HS026257-02 from the Agency for Healthcare Research and Quality

Grant Award Number.
R21 HS026257-01

Structured Abstract

Purpose: Compare the performance of preventive measures ranked using a previously piloted longevity-estimating algorithm for Individualizing Precision Prevention (IPP) with rankings from Primary Care Providers (PCPs).
Scope: One IPP algorithm compared for concordance with collective rankings from 40 PCPs.
Methods: Technology development followed by the use of a special ranking questionnaire with 12 realistic patient scenarios culminating in data collection and analysis involving Length-dependent Rank-biased Overlap (LDRBO) calculations.
Results: For all 12 scenarios, comparing the IPP algorithm to the combined rankings from all PCPs yields a mean value of .45, corresponding to a moderate level of concordance or agreement between the rankings of the IPP algorithm and the provider rankings.
Key Words: Prioritization, Preventive Medicine, Informatics, USPSTF, Individualization

Purpose and Objectives of Study

Purpose

Since 1984, the United States Preventive Services Task Force (USPSTF) has conducted numerous systematic reviews of the evidence for multiple preventive services. In this manner, the USPSTF determines the level of scientific evidence supporting a preventive care recommendation, assigning well-supported recommendations evidence grades of A or B[1].

Currently, USPSTF guidelines recommend about 50 preventive care services based on grade A or B evidence. However, the US health system has yet to maximize the benefit of these preventive care recommendations. Problems include overutilization and underutilization of recommended preventive care services and large subpopulation disparities in preventive care and related health outcomes. Currently, the number of recommended preventive services is more than can be consistently applied in primary care and is still growing. To optimize social benefit, capabilities to prioritize preventive services precisely are needed.

The idea of the Learning Health System (LHS), which we think of as both an inspiring concept and a systematic and scalable set of methods, has gained increasing recognition since the appearance of a seminal report from the Institute of Medicine[3]. The LHS “learns” through a cyclical process that engages an interested community in the assembly and analysis of data relevant to an important problem, which leads to the discovery of new knowledge from the data[5]. The learning cycle is completed by direct application of that knowledge to change practice, typically making use of recommendations individualized to the characteristics of each patient and tailored to the needs of care providers. In this project, we take advantage of the existing USPSTF recommendations, which constitute high-quality knowledge, and focus on enhancing the knowledge-to-practice component of the learning cycle to facilitate personalized, prioritized uptake of these recommendations to optimize health.

More specifically, we apply LHS methods to advance toward effective Individualized Precision Prevention (IPP) for non-pregnant adults. The IPP approach both personalizes and prioritizes preventive services systematically and scalably. This work is supported by the Knowledge Grid (KGrid), a broadly applicable technical platform created at the University of Michigan to support the knowledge-to-practice aspect of LHS[4]. KGrid includes Knowledge Objects, a digital Library within which to hold and manage them, and an Activator with which to deploy them as digital services. Enhancing this shared infrastructure, the Substitutable Medical Applications and Reusable Technologies (SMART) initiative supports an app ecosystem to extend the capabilities of EHRs[5]. For this project, we used KGrid to build and test a new web application capable of automatically personalizing and prioritizing preventive measures to achieve IPP.

The main purpose of this study is to address the challenge of providing evidence-based IPP. To achieve that purpose, we demonstrate how KGrid, when used to manage and deploy computable knowledge, and open source web application technologies can work together to quickly generate IPP information in the form of patient-specific, rank-ordered lists of recommended preventive services based on evident relative benefits for health. The results of the project include sharable computable USPSTF knowledge objects, a shareable web app, and initial insights into how to effectively use these resources to individualize preventive services.

Objectives

Objective 1

Use KGrid to design and develop computable knowledge objects for USPSTF A & B recommendations, along with an executive object capable of applying computable risk and benefit models, forming a core knowledge object collection for Individualized Precision Prevention (KOs-4-IPP).

Objective 2

Design, develop, and test a shareable SMART IPP APPLICATION that can be integrated with various EHRs and which draws on KOs-4-IPP services from KGrid to provide individualized precision prevention information.

Objective 3

Conduct a study with primary care providers as subjects to assess the potential utility of IPP information.

Scope and Methods of Study

Scope

Overview and General Context.

The practice context for this work is the primary care environment where decisions are made about which preventive medical services to offer to individual patients. We prepared technology and a system capable of supporting individual precision prevention (IPP) and then performed a content validation and utility study of the concordance of the IPP system's output with providers' self-reported current approaches to implementing evidence-based preventive measures.

Setting and Participants.

For the study of concordance, we recruited 40 PCP attendees at a regional, professional meeting, each with five or more years of clinical practice experience. Each participant was randomly assigned to one of the four sets containing 12 scenarios and was given a list of the USPSTF A and B recommendations for which the patient described in each scenario is eligible. Each subject worked through their randomly assigned set of 12 scenarios in a random sequence to limit order effects. Our experimental protocol had three steps. First, after reading a scenario, each subject was provided a list of all of the preventive services for which the individual described by the scenario is eligible. They were asked to rank the relative importance of each service, in descending order of importance. Second, because there may be differences between PCP perceptions of importance and how they would decide to address preventive services in practice, we asked subjects to identify and confirm the one preventive service they would choose to discuss first with each hypothetical patient if time was limited as during a routine clinic visit. Third, subjects filled out a questionnaire with several open-ended questions. The open questions allowed PCPs to share their rationale for ranking preventive services (by importance) as they did.

Methods

Approach for Aim 1

Our approach to achieving Aim 1 involved technical design and software development work. Supporting this effort is our prior work on the Knowledge Grid (KGrid) technical infrastructure program over the past several years. From 2016-2018, we designed, developed, and formally modeled digital *knowledge objects* (KOs) and published a paper on the conceptual design of KOs.

KOs are packaged compound digital objects used for externalizing and modularizing computer logic for shareable, interoperable clinical decision support. KOs have three main parts. First and foremost, KOs hold a *computable biomedical knowledge payload*. The computer-processable payload could be expressed using executable rules, or a mathematical function expressing the results of empirical analysis (e.g., a regression equation), or as a table of known values (e.g., population statistics from CDC). Second, KOs have a *service description*. The service description part of KOs contains information to tell people and machines how to interact with the computable biomedical knowledge payload held in the KO. For this project, the service description specifies an application programming interface (API) for each KO. Third, KOs include metadata describing their content, origin, lifecycle, etc. These metadata are important for making KOs findable, accessible, interoperable, and reusable (the FAIR principles).

Prior to starting this project, we developed 70 KOs of different types, including some KOs holding statistical risk models and others holding computable guidelines. For this project, we used previously tested technical methods to build the "KOs-4-IPP" collection. These technical methods included encoding USPSTF A and B recommendation information, mathematical population health functions, and tables of public health statistics using JavaScript; describing KO APIs in service descriptions using the OpenAPI 3.0 standard, storing KOs in GitHub, and packaging KOs as ZIP files. The new KOs-4-IPP collection essentially modularized computable knowledge using KOs, allowing that knowledge to be curated, combined, deployed, and updated by our study team using KGrid technology. More information about KGrid technology is provided in Appendix C.

As part of the KOs-4-IPP collection, we designed and developed an Executive KO for the first time. This KO plays a special role in the KGrid technical infrastructure. By design, Executive KOs have computer-processable payloads that orchestrate computations using other KOs. For this project, we first upgraded the KGrid Activator to support the capabilities of Executive Objects with JavaScript payloads. The KGrid Activator is a Java microservice built using Spring tooling. The Activator encapsulates or otherwise integrates various programming-language runtimes, including, in this case, both the existing Nashorn and V8 JavaScript runtimes. Once the Activator was suitably upgraded, the first basic Executive KO with a JavaScript payload was designed and tested. From there, we designed and tested a more sophisticated Executive KO to meet the needs of this project.

Approach for Aim 2

Our activities to achieve Aim 2 focused on developing an easy-to-use software app. Our original intent was to develop a SMART app. However, at the time of our app development for this project, we found most of the inputs we needed for an IPP were not available to us as FHIR resources from the EHR platforms of interest. For this reason, we pivoted to create an IPP web app and are now returning to work on a FHIR-based IPP app in 2020.

For this project, our web app development work unfolded in two phases. The first phase was for design and development. The second phase was for testing. Using iterative, agile, test-driven software application development methods, we designed and built an IPP web app that runs in common web browsers. Our IPP web app is a JavaScript app built using the Vue framework.

The IPP web app computes with a specified list of input data elements about an individual. The app relies on the KOs-4-IPP collection and engages the Executive Object from the collection to do all of its individualized precision prevention computations. We deployed the IPP web app on a common webserver, from where it can be accessed online.

To test the IPP web app, we used synthetic test patient data provided by Dr. Glen Taksler of the Cleveland Clinic. For these data, the expected IPP outputs were previously known. Dr. Taksler and Dr. Caverly reviewed the outputs computed by our IPP web app and confirmed that the outputs were appropriate and in line with their combined experience and expectations.

Approach for Aim 3

Goal of the Study:

To evaluate the potential importance of having computed IPP priorities readily available for point-of-care decision-making by primary care providers (PCPs), we conducted a laboratory experiment, using realistic patient scenarios, to test the ***concordance*** between the computed IPP priorities and experienced PCPs' perceptions about the relative importance of specific preventive services.

We hypothesized that a very high concordance would indicate that the computed priorities, offered as advice, would not add novel information in practice, although it still might be useful to confirm clinicians' own judgments. Moderate concordance would suggest that the computed priorities an opportunity to improve the effects of preventive care. Low concordance would be of some concern, calling into serious question either the validity of the algorithm or that of clinicians' judgments about the preventive services of most potential benefit to their patients.

Overall Study Design:

We systematically developed 12 patient scenarios, spanning a range of health statuses and a range of eligibilities for preventive services. We developed a data collection instrument to be completed by primary care practitioners (PCPs) under proctored conditions. The instrument included all twelve scenarios and asked practitioners, after reading each scenario, to order in priority those preventive services for which the patient described in the scenario was eligible. The scenarios as, presented in the instrument, included all of the clinical data used by the Taksler algorithm[6]. The instrument was created in two similar versions with just one key difference. The "Green" version asked for priority rankings based on ***extending the patient's life*** and the "Blue" version asked for rankings ***in favor of maintaining the patient's overall health***. Because the survey was NOT intended to test subjects' knowledge of USPSTF recommended preventive services, the survey clearly indicated which USPSTF recommended preventive services each simulated patient was eligible for. Independently, we computed the priority of each of the preventive services using the Taksler algorithm[6] as implemented in the Knowledge Grid. We administered the survey to 40 PCPs, half randomly assigned to each version and with each PCP encountering the scenarios in random order. We then compared statistically the rankings offered by the PCPs with those computed by the Taksler algorithm[6].

Patient Scenarios:

Collectively, the set of 12 scenarios spanned the spectrum of preventive health services through systematically engendered variance across demographics, age, health habits, overall health status, and risk profiles. Each scenario script included the structured data necessary to compute all applicable Grade A and B USPSTF recommendations as well as the priorities assigned to each recommendation via the Taksler algorithm[6]. Augmenting the structured data, we wrote portrayals, in prose, describing each of the “patients”, taking care not to add extraneous clinical details that could influence clinician decisions about preventive care services.

We used the American Society of Anesthesiologists (ASA) health status classes as the primary stratification factor[7]. ASA Class I represents people who are healthy non-smokers. ASA Class II represents mild diseases only without substantive functional limitations, such as current smokers, and those with controlled diabetes or hypertension. ASA Class III is for those with poorly controlled disease (in our scenarios: those with poorly controlled diabetes or high blood pressure and who are morbidly obese with a Body Mass Index > 40 and who have experienced a prior heart attack). We created 4 “health status” scenarios in each ASA Class--2 males and 2 females--and we identified the preventive services for which each “patient” was eligible. One of the 12 health status scenarios is displayed in Figure 1 below:

Health status: ASA Class III
59-year-old woman:
BMI=46; poorly controlled DM and HTN; current smoker
History of prior heart attack with stents placed 3 years ago

She is eligible for the following USPSTF recommended preventive services:
Cervical cancer screening; CRC screening; HIV screening; HTN screening; Alcohol misuse; Tobacco cessation; Aspirin use; BRCA-related cancer risk assessment; Breast cancer preventive medications (if high risk); Breast cancer screening; Chlamydia and Gonorrhea screening; Depression screening; Healthful diet & activity counseling; Hepatitis C screening; Lung cancer screening; Osteoporosis screening (if high risk), Sexually transmitted infections counseling; Statin use; Weight loss interventions; Skin cancer counseling

Figure 1: An example health status scenario.

We then embellished each of the 12 health status scenarios, as shown above, with more complete clinical information consistent with the patient’s overall health status. These more detailed descriptions contained all information required to compute, using the Taksler algorithm, the priority of each service. These data were added based on the clinical judgment of an experienced general internist who was not aware of how the algorithm would rank these eligible preventive services.

The Data Collection Instrument:

We developed a data collection instrument from the patient scenarios. For each fictitious but highly-realistic patient scenario, the data collection instrument has two parts shown in Figures 2 and 3. An example of the first part of our data collection instrument describing the scenario for one of our “patients” is shown in Figure 2 below.

Patient Demographics		
Age	Gender	Race / Ethnicity
79	Female	Caucasian

Current Medications

Lisinopril 40mg PO QAM
Amlodipine 5mg PO QAM
Metformin 1,000mg PO BID
ASA 81mg PO QDAY

Active Problems

HTN (poorly-controlled; averages 160s/90s at home)
Type II Diabetes (poorly-controlled; last A1c 8.6%)
CAD with prior MI and stenting

Social History

Tobacco Use	*Years Used*	*Avg. Packs/Day*	*Pack-years*	*Years Since Quit*
Current Smoker	60	1	60	0

Alcohol consumption

Risky use (averages 3 drinks per day; occasional binge drinking)

Sexual History

+history of sexually transmitted disease

Family History

Sister with pancreatic cancer
Father died following an MI
Mother with previous fragility fracture

Reproductive history and bone health

Menarche at age 13
Menopause at age 46
1 child at age 28
Previous forearm fracture, no parental history of hip fracture, no medication use or medical history associated with osteoporosis
HIGH fracture risk based on FRAX score

Patient data	
Height	67"
Weight	255 lbs
Body Mass Index	40 (Morbidly obese)
Blood Pressure	162/90
Total Cholesterol	280
HDL	32
LDL	178
Triglycerides	350
10-year CVD risk	75.0% (ASCVD)
Serum Creatinine	0.9 mg/dL

Physical Activity

Does not achieve recommended exercise guidelines from the Department of Health and Human Services.

Figure 2: Example Patient Scenario from Data Collection Instrument

The second part of our data collection instrument provides a study participant response form listing all of the USPSTF “A and B” preventive services for which the patient described in the scenario is eligible. This form is used to collect the study subject’s ranking of their “Top 5” preventive services to be implemented for the patient in practice, as illustrated in Figure 3 below. Subjects were also asked to rate their confidence in their Top 5 rankings.

Your Preventive Services Ranking for Ms. L.
For five of the preventive services in the list below, respond with a number corresponding to the rank order of that service for its importance to maintaining this patient's overall health. Because ties are not permitted in your "top 5" please do not use any number more than once. For the remainder of the services, indicate either that the service is not in your "top 5" in priority for this patient OR that you would not recommend this service regardless of its priority for this patient. Please give a response for all listed services and note that the services are listed in alphabetical order.

	Highest						
Alcohol use: reduce to healthy levels	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Blood pressure: add or intensify medication	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Depression screening	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Diabetes: add or intensify medications	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Healthful diet & activity counseling	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Lung cancer screening	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Osteoporosis screening	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Stop smoking	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all
Weight: lose 10 lbs	1	2	3	4	5	Not in My Top 5	I would not recommend this service at all

How confident are you in the rankings you have given above?

Completely Confident	Fairly Confident	Somewhat Confident	Slightly Confident	Not Confident At all
1 ○	2 ○	3 ○	4 ○	5 ○

Figure 3: Priority and Confidence Elicitation

Before using our 12 case scenarios for our principal data collection, we performed a pilot study with three PCPs to get feedback on the clarity of our instructions. Minor changes to the content and the format of the instrument were made based on their questions and suggestions.

<u>Procedure</u>:

Before beginning our study, we created 40 packets, each one containing a randomly ordered set of the 12 scenarios and preventive services lists. We put the scenarios in a random order to mitigate order effects knowing that our subjects would work through each packet scenario-by-scenario in a serial fashion. Half of the packets, randomly assigned, asked for priority rankings based on extending the patient's life and the other half asked for rankings in favor of maintaining the patient's overall health.

We recruited 40 PCPs from the State of Michigan, the majority of whom participated in our study during a Family Medicine conference held in Ann Arbor, Michigan on October 2-3 2019. A few of the study participants were recruited from the greater Ann Arbor area after the conference. For data collection, each participant was randomly issued one of the experimental packets with 12 scenarios and corresponding USPSTF A & B preventive service recommendation lists.

Analytical Plan:

To compute the relationship between the providers' priorities and those generated by the algorithm, we employed the Length-dependent Rank-biased Overlap (LDRBO) statistic developed by co-investigator Philip Boonstra[8]. This statistic is necessary to support this analysis because traditional measures of correlation, such as Spearman's ρ or Kendall's τ, are inappropriate for ranking partial lists of unequal length. The LDRBO statistic is defined in more detail in Appendix B.

We compared the providers' and the algorithm's rankings by, first, aggregating the 40 providers' lists of recommendations into a single derived 'consensus' list of recommendations, defined as the hypothetical list of exactly three recommendations that maximizes the median LDRBO similarity across the 40 providers. To reflect the reality of a clinical setting in which a provider can typically make and enact at most three recommendations in a visit, we considered only the top three rankings from the algorithm and each provider. We then computed the LDRBO statistic relating the algorithm's rankings to the consensus rankings of the providers. We conducted the analysis separately for the forms using the "green" and "blue" ranking criteria as described in the methods section above, and also for both form versions combined.

Results

Results for Aim 1

Overall, the study team successfully completed the needed technical work to achieve Aim 1. The KOs-4-IPP collection of Knowledge Objects includes 1 Executive KO, 1 Recommendation List KO, 1 Life Table KO, 2 Total Background Risk KOs for figuring population-level risks, 11 Patient Background Risk KOs for scoring individual patient risks (e.g., risk of cardiovascular disease), 17 USPSTF Recommendation Net Benefit Calculator KOs, 7 Patient-Derived Feature KOs, 1 Life Expectancy calculator KO, and 1 Life Expectancy *Gain* KO. These 42 KOs comprise the KOs-4-IPP collection and are described in more detail in Appendix A.

For these results, two KOs out of the 42 in the KOs-4-IPP collection warrant further description. The first of these is the IPP project's new Executive KO. Because the Executive Object was designed and developed first for this project, its design and development are significant project results in their own right. The key to our JavaScript-based Executive Objects' capabilities is to have a JavaScript function that calls other KO API endpoints after those endpoints have been established in the *same running instance* of the KGrid Activator. The function inside the Executive Object payload that does is shown in Figure 4 below.

```
function execute(endpoint, inputs){
  return context.getExecutor(endpoint).execute(inputs)
}
```

Figure 4: JavaScript function supporting the capability to have Executive Objects

Note how the JavaScript function above accepts and used two parameters, one called *endpoint* and the other called *inputs*. Using the KGrid Activator integrated with a JavaScript runtime, this function accepts and processes a locally-available API endpoint (e.g., `ipp-bmicalculator/v0.0.3/bmi`) and appropriate inputs for engaging the endpoint to compute (e.g., height and weight). Within the scope of the Executive Object, the example above of a call to the execute function would return a computation of a patient's body mass index (BMI). Thus, the execute function, working in conjunction with the Activator, enables Executive KOs to perform their executive role by engaging APIs that arise from other KOs.

The second KO of interest to further describe in these results is the USPSTF Recommendation List KO. This KO uses facts about an individual to determine which USPSTF Recommendations apply to that individual's care. Inside this KO we declare inclusion or exclusion criteria for each USPSTF Recommendation covered by the IPP-4-KOs collection. Below in Figure 5 is one example of this.

```
"aaascreening":{
    "name": "Abdominal Aortic Aneurysm: Screening",
    "shortText":"Abdominal Aortic Aneurysm: Screening",
    "description":"The USPSTF recommends one-time screening for abdominal aortic aneurysm (AAA) with
ultrasonography in men ages 65 to 75 years who have ever smoked. ",
    "type": "Screening",
    "releaseDate": "June 2014",
    "grade":"B",

"uspstflink":"https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/abdominal-
aortic-aneurysm-screening",
    "basePopulation": {
      "population":"Male",
      "minimumAge":"65",
      "maximumAge":"75"
    },
    "benefitko":"ipp-aaascreening/v0.0.3/netbenefit"

  },
```

Figure 5: Encoded inclusion and exclusion criteria for the Abdominal Aortic Aneurysm Screening Preventive Service Recommendation

The example in Figure 5 covers the inclusion criteria for abdominal aortic aneurysm screening. At the moment with the KOs-4-IPP collection was being developed, this recommendation pertained only to men between the ages of 65 and 75 years. For that reason, we encoded these inclusion criteria so that, with sex and age inputs, we can determine whether or not this recommendation applies to any adult.

The other KOs in the KO-4-IPP collection are all marshaled by the Executive KO and needed to compute relevant USPSTF recommendations and coinciding individualized prevention services priorities based on life-gain estimates for any adult.

Results for Aim 2

The IPP web app we created to test and trial the KOs-4-IPP collection is deployed online here: https://kgrid-objects.github.io/ipp-collection/web/#/ .

The user interface for the IPP web app supports a single patient process of prioritizing relevant USPSTF A & B recommendations. As portrayed in Figure 6 below, a panel of data elements about an individual patient characteristics arrayed on the left are used to compute the relevance and utility of USPSTF A&B recommendations listed on the right. To do these computations, the KOs-4-IPP collection is instantiated, made runnable, and engaged by the web app using the KGrid Activator.

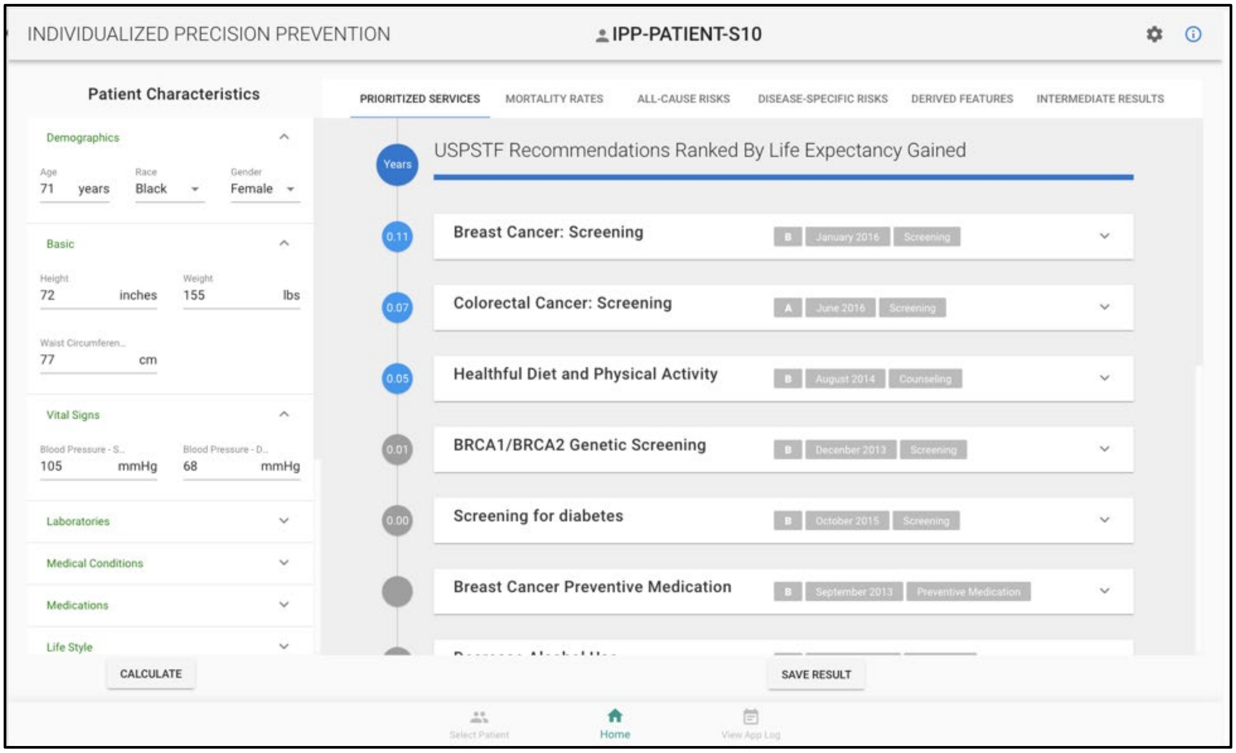


Figure 6: User View of IPP web app showing patient characteristics on the left and ranked relevant preventive services on the right

In the example portrayed in Figure 6, the IPP web app has ranked relevant USPSTF A & B recommendations for a fictitious 71-year-old black female who is 6 feet tall, weighs 155 pounds, and has low blood pressure (105/68). In her case, the greatest life-gain value estimated (0.11 years) would come from implementing a USPSTF A & B recommendation for Breast Cancer Screening. Colorectal cancer screening and a healthy diet are also estimated to bring modest gains in her longevity. More results like these computed for individuals are featured in the results for Aim 3, which come next.

Results for Aim 3

Table 1, below, provides the LDRBO-based comparison of ordered lists arising from the algorithm versus a consensus-based aggregation of the 40 providers' ranked lists, combined and separately for the "Green" and "Blue" ranking criteria. Results are presented separately for each of twelve scenarios, arranged by decreasing values of LDRBO. A legend to decipher the codes for unique preventive services in Table 1 can be found immediately below in Table 2.

Scenario	Source of Ranking	Top-ranked Service	Second-ranked Service	Third-ranked Service	Value of LDBRO	Number of Shared Services in Top 3	**Scenario**	Source of Ranking	Top-ranked Service	Second-ranked Service	Third-ranked Service	Value of LDBRO	Number of Shared Services in Top 3
4	Algorithm	BRE	CRC	DIE			**2**	Algorithm	LUN	STA	DIE		
	Consensus (ALL)	BRE	CRC	CER	0.89	2		Consensus (ALL)	STA	DIE	CRC	0.39	2
	Consensus (GREEN)	BRE	CRC	CER	0.89	2		Consensus (GREEN)	STA	DIE	ASA	0.39	2
	Consensus (BLUE)	BRE	CRC	CER	0.89	2		Consensus (BLUE)	STA	DIE	CRC	0.39	2
6	Algorithm	BP	SMO	ALC			**7**	Algorithm	CRC	DIE	AAA		
	Consensus (ALL)	SMO	BP	ALC	0.67	3		Consensus (ALL)	DIE	STA	CRC	0.39	2
	Consensus (GREEN)	SMO	BP	DIE	0.56	2		Consensus (GREEN)	DIE	STA	CRC	0.39	2
	Consensus (BLUE)	BP	DIA	BRE	0.61	2		Consensus (BLUE)	DIE	STA	AAA	0.39	2
5	Algorithm	SMO	BP	DIE			**8**	Algorithm	DIE	WEI	LUN		
	Consensus (ALL)	SMO	BRE	CER	0.61	1		Consensus (ALL)	STA	DIE	WEI	0.39	2
	Consensus (GREEN)	SMO	BRE	CER	0.61	1		Consensus (GREEN)	STA	DIE	WEI	0.39	2
	Consensus (BLUE)	SMO	BRE	DIE	0.72	2		Consensus (BLUE)	STA	WEI	DIE	0.39	2
11	Algorithm	SMO	DIE	BP			**10**	Algorithm	BRE	CRC	DIE		
	Consensus (ALL)	SMO	CRC	STA	0.61	1		Consensus (ALL)	OST	DIE	BRE	0.22	2
	Consensus (GREEN)	SMO	OST	STA	0.61	1		Consensus (GREEN)	OST	DIE	CRC	0.22	2
	Consensus (BLUE)	SMO	DIE	STA	0.89	1		Consensus (BLUE)	OST	DIE	BRE	0.22	2
1	Algorithm	DIE	CRC	ASA			**12**	Algorithm	DIE	ALC	SMO		
	Consensus (ALL)	CRC	DIE	DEP	0.56	2		Consensus (ALL)	BP	SMO	ALC	0.22	1
	Consensus (GREEN)	CRC	DIE	DEP	0.56	2		Consensus (GREEN)	BP	SMO	DIA	0.11	1
	Consensus (BLUE)	CRC	DIE	DEP	0.56	2		Consensus (BLUE)	BP	DIA	DIE	0.11	0
3	Algorithm	ALC	DIA	BP			**9**	Algorithm	ALC	DIE	WEI		
	Consensus (ALL)	BP	DIA	ALC	0.5	3		Consensus (ALL)	BP	STA	DIA	0	0
	Consensus (GREEN)	BP	DIA	STA	0.39	2		Consensus (GREEN)	BP	STA	DIA	0	0
	Consensus (BLUE)	BP	DIA	DIE	0.39	2		Consensus (BLUE)	BP	STA	ALC	0.11	0

Table 1: Comparison of algorithm-ranked preventive services with consensus rankings of the same preventive services by providers. For each of our 12 numbered scenarios and their fictitious patient cases, concordance is indicated by the value of LDBRO (higher values show more concordance) and by the number shared preventive services in the Top 3 as ranked by the IPP algorithm and by the providers. For each scenario, first the Top 3 services ranked by the algorithm are given. Next, the consensus of the algorithm and ALL of the providers is given. Finally, the consensus of the algorithm the GREEN provider group and BLUE provider group are given.

To interpret the information above in Table 1, here we provide a second table as a legend for the preventive services represented by the 16 different three-letter codes in the table.

AAA	Abdominal aortic aneurysm screening
ALC	Alcohol use: reduce to healthy levels
ASA	Aspirin use
BP	Blood pressure: add or intensify medication
BRE	Breast cancer screening
CER	Cervical cancer screening (last screen 3 years ago; no prior abnormal screens)
CRC	Colorectal cancer screening
DEP	Depression screening
DIA	Diabetes: add or intensify medication
DIE	Healthful diet & activity counseling
HIV	HIV Screening
OST	Osteoporosis screening
LUN	Lung cancer screening
STA	Statin use
SMO	Stop smoking
WEI	Weight: lose 10 lbs.

Table 2: Legend for three-letter preventive services codes showing the codes and the services they represent in this study

Averaging the LDRBO values shown in Table 1 above for all 12 scenarios yields a mean value of .45, corresponding to a moderate level of concordance or agreement between the rankings of the IPP algorithm and the provider rankings. The specific points of agreement and diversion can be seen by comparing the specific preventive services that would be in the "Top 3" by provider consensus and the services that would be in the algorithm's "Top 3".

The rightmost column of Table 1 presents the number of preventive services that appear in both the providers' consensus list and in algorithm's list. A value of 3 in this column represents situations where the algorithm and providers recommended the same "Top 3" preventive medical services even if the Top 3 services are ranked in a different order. Looking across all providers, for 6 of the 12 cases, the algorithm would have recommended one service that was not in the providers' consensus Top 3; in 3 of 12 cases, the algorithm would have recommended 2 services that were not in the providers' Top 3.

Discussion, Implications, Limitations, and Future Work

Pursuant to Aims 1 and 2

From a technical perspective, using maturing approaches we have pioneered to modularize, package, and combine instances of computable biomedical knowledge, we have demonstrated advancement by achieving the first aim of this project. Using a moderately-sized collection of Knowledge Objects, some containing payloads with computable inclusion and exclusion criteria for USPSTF A and B recommendations and others containing information and computational know-how to estimate the life-gain from each recommendation for individuals, we instantiated and used a complex IPP algorithm. Moreover, because our technical approach externalizes computable biomedical knowledge using shareable Knowledge Objects, the IPP algorithm is not embedded inside a single end-user application. Rather, our implementation of this algorithm using the KOs-4-IPP collection stands on its own. This collection can be implemented as a straightforward webservice with a RESTful API and accessed by any end-user application of interest that way. This is a small breakthrough in making complex, multi-part algorithms comprised of computable biomedical knowledge more easily transferable and more widely available.

One key technical advancement for the Knowledge Grid program achieved through this project is the development and demonstration of the first Executive Knowledge Object. This important step paves the way for future implementations of complex biomedical models comprised of many submodels and other information resource subcomponents. Unlike traditional monolithic clinical decision support systems (CDSS), our approach is to manage complex, multi-component instances of related computable biomedical knowledge as resources and services in their own right. Achieving a functional capability to have and use Executive Knowledge Objects is critical for externalizing and managing computable biomedical knowledge separately from CDSS.

We have several limitations to report. In our technical work, we encountered barriers specifically with the availability and representation of the approximately 100 inputs needed to compute using the IPP algorithm instantiated in the KOs-4-IPP collection of Knowledge Objects. Specifically, we found that a wide variety of the needed inputs other than demographics and lab values were not yet commonly represented using HL7 FHIR resources. We expect that this will change as more and different types of FHIR resources become available. For the time being, we developed our own proprietary input schema recognizing that it is insufficient and in need of improvement.

Pursuant to Aim 3

The LRDBO values relating clinicians' rankings of preventive services to those generated by the Taksler algorithm fell into the anticipated "middle range". The magnitude of this relationship strikes a balance between irrelevance, in which case the clinical applicability of the algorithm's rankings would be subject to doubt, and redundancy, in which case the algorithm's rankings would be superfluous. For half of the 12 cases, the algorithm would have recommended one preventive service that was not in the providers' consensus Top 3.

We were, at the same time, impressed by the high level of case-to-case variability in the LRBDO values and plan to study, as an unfunded extension of the originally proposed work, to develop and publish a statistical model relating LRDBO to the clinical data in each of the case scenarios. This analysis will shed light on the characteristics of patients for which providers' subjective sense of what preventive services might be appropriate for that patient varies from the objective recommendations of the algorithm. We also plan to extend the study focused on ranking concordance by computing the LRDBO value by subject as an alternative analytical strategy to the consensus approach reported above.

Limitations

This study is limited by the breadth and size of the provider sample. These 40 individuals were from one state and had, with a small number of exceptions, self-selected to attend a family medicine continuing education conference. This convenience sample is not a representative sample of PCPs throughout the nation. Also, a sample larger than forty would result in a more precise estimate of the consensus preventive services rankings, and the sources of variability around the consensus, than our limited sample size could provide. An additional important limitation to the study goes to the external validity of a laboratory study, as opposed to a clinical field study, using hypothetical case scenarios, as opposed to actual patients seen in clinical environments.

Future Work

Finally, our work on IPP now extends beyond the three aims of this study. We are pursuing an updated IPP algorithm capable of accounting for quality of life as well as longevity. Our current plans are to deploy an IPP app in the EHR for a clinical pilot at Cleveland Clinic in early 2022.

Conclusion

Here we report success developing and implementing a complex IPP algorithm to rank order USPSTF A and B level recommendations for preventive services for non-pregnant adults. We leveraged technical advancements in computable biomedical knowledge management capabilities enabling us to externalize and combine multiple instances of computable biomedical knowledge to implement the IPP algorithm. Using realistic scenarios, when testing concordance between the IPP algorithm's ranking of preventive service recommendations and provider rankings of those same services, we found an intermediate level of concordance. This finding suggests that there may be a role for IPP algorithms to help guide selection and enactment of preventive services in practice.

References

1. US Preventive Services Task Force, United States. Office of Disease Prevention, Health Promotion. Guide to clinical preventive services: report of the US Preventive Services Task Force. US Department of Health and Human Services, Office of Public Health and Science, Office of Disease Prevention and Health Promotion; 2014.
2. McGinnis JM, Aisner D, Olsen L, editors. The learning healthcare system: workshop summary. National Academies Press; 2007 Jun 1.
3. Friedman CP, Rubin JC, Sullivan KJ. Toward an information infrastructure for global health improvement. Yearbook of medical informatics. 2017 Aug;26(1):16.
4. www.kgrid.org, accessed December 6, 2020.
5. Mandl KD, Mandel JC, Murphy SN, Bernstam EV, Ramoni RL, Kreda DA, McCoy JM, Adida B, Kohane IS. The SMART Platform: early experience enabling substitutable applications for electronic health records. Journal of the American Medical Informatics Association. 2012 Jul 1;19(4):597-603.
6. Taksler GB, Keshner M, Fagerlin A, Hajizadeh N, Braithwaite RS. Personalized estimates of benefit from preventive care guidelines: a proof of concept. Annals of internal medicine. 2013 Aug 6;159(3):161-8.
7. Doyle DJ, Garmon EH. American Society of Anesthesiologists classification (ASA class). InStatPearls [Internet] 2019 May 13. StatPearls Publishing.

8. Krauss JC, Boonstra PS, Vantsevich AV, Friedman CP. Is the problem list in the eye of the beholder? An exploration of consistency across physicians. Journal of the American Medical Informatics Association. 2016 Sep 1;23(5):859-65.

List of Publications and Products

Planned Publications

We plan to publish two papers based on the results in this report. The first paper will target a technical informatics audience and discuss how the complex IPP algorithm was implemented successful using a modularized approach founded on Knowledge Objects. The second paper will target a health services research and clinician audience and focus on the results and the implications of partial concordance between the IPP algorithm's ranking of preventive medical services and the collective ranking of providers.

Online Research Products Produced by This Study

The following products can be accessed simply by making a request to the following email address: kgrid-developers@umich.edu

(1) Individual Precision Prevention Collection of Knowledge Objects stored here: https://github.com/kgrid-objects

(2) Individual Precision Prevention Collection Web Application stored here: https://github.com/kgrid-demos

(3) Core Knowledge Grid Components (e.g., latest Activator) described here: https://demo.kgrid.org/

Appendix A

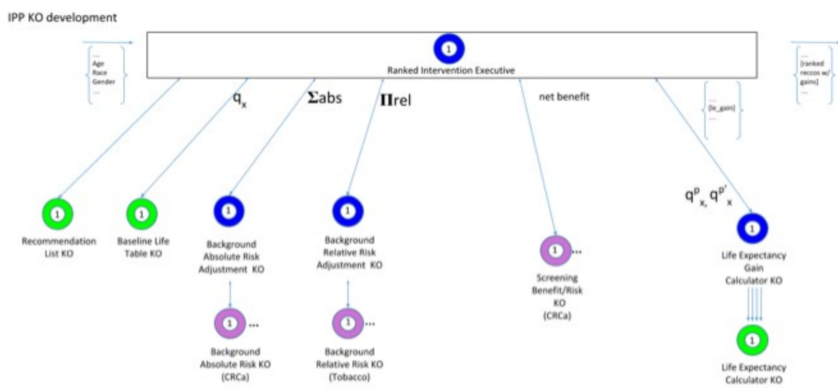
Individualized Precision Prevention Knowledge Objects

This project uses a complex, multi-part mathematical model to rank order preventive medicine measures that are recommended by the United States Preventive Services Task Force. It provides an individual rank order of these measures for each person, based on 97 person-specific features including height, weight, cholesterol, smoking status, etc.

The complex model has been implemented using Knowledge Grid technology. The model is highly modularized utilizing a collection of 42 IPP Knowledge Objects. Eight different Knowledge Object schemas (or types) comprise the IPP collection. These schemas are listed below, starting with the IPP Executive Object schema. The IPP Executive Object is interesting because it is an example of a Knowledge Object that calls on other Knowledge Objects to complete a complex series of calculations.

IPP Knowledge Object Support Computational Flow

The following depicts an example of one of the IPP processing workflows employed in this collection.



IPP Knowledge Objects (KO)

The following schemas are designed for this project and are referenced in the service description YAML file using `$ref`.

- Patient Feature Panel

A sample patient feature panel is used by the demo app and can be found at:

- IPP Patient Features Sample

The collection consists of the following types:

IPP Executive Object *(1 Knowledge Object)*

The executive KO coordinates the workflow for evaluating the risk profile, computing additional derived features, retrieving mortality rate, computing life expectancies and life expectancy gains per USPSTF recommendations and generating a ranked list of the preventive medicine recommendations.
The executive KO calls multiple KOs in several stages, aggregates the results, passes the data along and assembles the final output containing the ranked recommendation list.

Recommendation List Knowledge Object *(1 Knowledge Object)*

The recommendation list KO contains a list of USPSTF A/B recommendations and takes a patient features panel as input to perform the following functions:

- Checks the patient features against `basePopulation` to determine if the recommendation is applicable for a patient
- Returns a map of all applicable Recommendations
- In the response, the relevant KO endpoints are presented in `benefitko` for the downstream computation

Life Table Knowledge Object *(1 Knowledge Object)*

The life table KO contains the data published by the CDC and it returns the mortality rate based on the patient's race and gender.

Patient Total Background Risk Knowledge Object *(2 Knowledge Objects)*

The Patient Total Background Risk KOs assess the risk profile based on the patient characteristics.
Based on Dr. Taksler's model, two KOs are developed:

- A background relative risk KO: Computes the total all-cause mortality risks as a product of the risks from: Tobacco, Obesity and Alcohol;
- A background absolute risk KO: Computes the total disease-specific mortality risks as a sum from individual background risk KOs.

Each KO calls the relevant individual risk KOs, aggregates the risks and returns the results to the executive KO.
Each KO also computes certain risks for the target characteristics as needed. For example, for "Decrease Alcohol Use", the risks need to be computed both for the patient's current drinking type and for the target type of "Abstain".

Patient Background Risk Knowledge Object *(11 Knowledge Objects)*

Each of The Patient Background Risk KOs assesses a particular risk associated with a condition based on the patient characteristics.
Based on Dr. Taksler's model, two types of background risks are computed:

- A background relative risk KO: Computes the all-cause mortality risks, including: Tobacco, Obesity and Alcohol;
- A background absolute risk KO: Computes the disease-specific mortality risks.

Each KO will compute the risk and return the results to the total risk KO.

USPSTF recommendation Net benefit Knowledge Object *(17 Knowledge Objects)*

Each Net Benefit Knowledge Object computes the net benefit by following the recommendation.
The result is returned to the executive KO.

Patient's derived feature Calculation Knowledge Object *(7 Knowledge Objects)*

A group of derived features are calculated in these KO and returned to the executive KO for later use.
This type of KO implements a well-known risk model, such as Framingham Risk score for CVD or CHD.

Life Expectancy Gain Calculation Knowledge Object *(1 Knowledge Object)*

This Life Expectancy (LE) Gain KO calculates the life expectancy gain based on the patient's risk-adjusted mortality rate set and the mortality rate set if the patient follows a USPSTF recommendation.
For each recommendation, The KO will call LE Calculator KO four times to compute: Total LE without screening/counseling Total LE with screening/counseling LE before next service without screening/counseling LE before next service with screening/counseling

Life Expectancy Calculator *(1 Knowledge Object)*

This Life expectancy calculator KO calculates the life expectancy based on a mortality rate set and the time span for valid contribution to LE.

For Total LE, the contribution from the current age till 100 will be included; while for the LE before next service, only the contribution from the current till next service time will be included. Next service time will be determined by the recommendation screening/counseling frequency per the recommendation.

Appendix B:

The Length-dependent rank-biased overlap (LDRBO) statistic

Given a pre-specified set of K items with integer labels $1, \dots, K$, an ordered list of these K items is any permutation of the integers, where the first integer in the permutation corresponds to the highest ranked item, the second integer to the second highest ranked item, and so forth. For comparing two ordered lists of all K items, classical measures of correlation like Spearman's ρ or Kendall's τ are appropriate. In contrast, these statistics are not suitable for comparing two partial lists, and they are not defined when comparing two partial lists of unequal length.

To that end, Boonstra proposed the LDRBO similarity measure. The LDRBO is itself an extension of the rank-biased overlap (RBO) of Webber designed specifically for lists of finite length. The LDRBO is a measure to compare the similarity of two ordered lists, even when they have unequal lengths.

Let $\mathbf{x} = \{x_1, x_2, \dots, x_{\ell_x}\}$ and $\mathbf{y} = \{y_1, y_2, \dots, y_{\ell_y}\}$ be two ordered lists of some library of items of size K, where ℓ_x and ℓ_y denote the length of $\mathbf{x}$ and $\mathbf{y}$, respectively, with $0 < \min(\ell_x, \ell_y) \leq \max(\ell_x, \ell_y) \leq K$. Given some prespecified positive-valued tuning parameter ψ, the LDRBO is defined as

$$\mathrm{LDRBO}_{\psi}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{d=1}^{\max\{\ell_x, \ell_y\}} \psi^d \, |\mathbf{x}_{1:d} \cap \mathbf{y}_{1:d}|/d}{\sum_{d=1}^{\max\{\ell_x, \ell_y\}} \psi^d}.$$

The expression $|\mathbf{x}_{1:d} \cap \mathbf{y}_{1:d}|$ denotes the size of the set-theoretic intersection of the first d elements of $\mathbf{x}$ and $\mathbf{y}$.

In words, the LDRBO is a number ranging between 0 and 1 defined as the weighted average of agreement across all possible depths d between the two lists, where 'agreement' is defined as the number of the first d elements common to both lists (or the entire list if d exceeds the list length) divided by d. The tuning parameter $\psi \in (0, \infty)$ controls the extent of rank-biasedness; *smaller* values correspond to *greater* ranked biasedness and *larger* values to *less* ranked biasedness. As the value of ψ is made closer to zero, the resultant value of LDRBO approaches 0 if the two lists disagree on their top ranked item or 1 if the two lists agree:

$$\mathrm{LDRBO}_{\psi \downarrow 0}(\mathbf{x}, \mathbf{y}) = 1_{[x_1 = y_1]}$$

Conversely, as ψ is made larger, the value of LDRBO approaches the number of items common to both lists divided by the longer of the of the two lists:

$$\mathrm{LDRBO}_{\psi \uparrow \infty}(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{\max\{\ell_x, \ell_y\}}.$$

In between, a value of $\psi = 1$ corresponds to a simple average across all agreements, and expression for LDRBO in this case reduces to

$$\mathrm{LDRBO}_{\psi=1}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{d=1}^{\max\{\ell_x, \ell_y\}} |\mathbf{x}_{1:d} \cap \mathbf{y}_{1:d}|/d}{\max\{\ell_x, \ell_y\}}.$$

In this document, we primarily use $\psi = 1$, such that when we use the generic term 'LDRBO' without an explicit value of ψ given, it should be read as $\mathrm{LDRBO}_{\psi=1}$.

Appendix C:

The Knowledge Grid and How it Works

The Knowledge Grid (KGrid) is an open-source platform for managing and running computable biomedical knowledge (CBK).

The kind of knowledge that works well in the KGrid might be risk calculators, computable guidelines, or reference and lookup tables — anything that can be represented as a set of services. A researcher or developer writes code to implement the knowledge as one or more functions, and the resulting code is packaged along with service and deployment descriptions as a knowledge object (KO).

The fundamental thing the Knowledge Grid does is allow you to externalize key pieces of computable biomedical knowledge that would otherwise be embedded in applications, EHRs, databases, and backend services. This makes it easier to reuse and update that knowledge, across time, for multiple channels, and in many organizations.

Basics

KGrid uses a "plugin" model. An activator component loads KOs at runtime, extracts and deploys the code to a suitable runtime environment, exposes the service the code implements as a simple RESTful API, and routes requests and responses. The service description (using OpenAPI 3) also specifies the inputs and outputs for the KO.

There is also a library component that can be used to manage and browse KOs. Since the activator and the library share a storage mechanism, they are typically deployed together. But one library can serve as a source of KOs for many activators, and one activator can import KOs from many libraries.

How it works

Currently, KGrid supports the embedded JavaScript engine, Nashorn, and a remote Node.js runtime. Additional runtimes are planned including an external Python environment, and cloud services like AWS Lambda and Google Cloud for serverless deployments. Knowledge objects are packaged as .zip files containing:

- a metadata file (metadata.json) containing identifiers and simple descriptive elements; the structural metadata follows the Knowledge Object Information Ontology (KOIO)
- code artifact(s)
- an OpenAPI .yaml document describing the service interface(s) the object provides
- a deployment descriptor .yaml document specifying the runtime environment(s), the entry point, etc.
- additional metadata if applicable

The activator and library are Spring Boot microservices written in Java. The library frontend is a Vue Single Page Application (SPA). They can be deployed directly in most environments. We also provide docker images for container scenarios.